DOCUMENT RESUME

ED 080 606 TM 003 125

AUTHOR	Bertram, Charles L.; And Others
TITLE	The Application of a Model for the Evaluation of Educational Products.
INSTITUTION	Appalachia Educational Lab., Charleston, W. Va.
PUB DATE	27 Feb 73
NOTE	109p.; Papers presented at symposium of annual meeting of American Educational Research Association (New Orleans, Louisiana, February 25-March 1, 1973)
EDRS PRICE	MF-$0.65 HC-$6.58
DESCRIPTORS	Career Education; *Decision Making; *Educational Programs; *Evaluation Methods; Intervention; *Models; Preschool Education; Program Development; *Program Evaluation; Speeches; Symposia

ABSTRACT

Papers presented at a symposium on "The Application of a Model for the Evaluation of Educational Products" are provided. The papers are: "A Model for the Evaluation of Educational Products" by Charles L. Bertram; "The Application of an Evaluation Model to a Preschool Intervention Program" by Brainard W. Hines; "An Evaluation Model for a Regional Educational Service Agency: Construction in a Product Development Setting" by Ermel Steep; and "The Application of the Evaluation Model to a Career Education Product" by Charles E. Kenoyer. In addition, a transcript of interaction among the symposium discussants, the panel, and the audience concerning the Model for Evaluation of Educational Products is presented. (For related documents, see TM 003 126-129.) (DB)

FILMED FROM BEST AVAILABLE COPY

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION
THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

ED 080606

The Application of a Model for the Evaluation of Educational Products

Charles L. Bertram
Gary Borich
Desmond L. Cook
Brainard W. Hines
Charles E. Kenoyer
Ermel Stepp

TM 003 125-129

Appalachia Educational Laboratory, Inc.

Charleston, West Virginia

P R O G R A M

THE APPLICATION OF A MODEL FOR THE EVALUATION OF EDUCATIONAL PRODUCTS

Chairman

Dr. John K. Hemphill, Executive Director, Far West Laboratory for Educational Research & Development

Participants

Dr. Charles L. Bertram, Director of Research and Evaluation, Appalachia Educational Laboratory
Mr. Brainard W. Hines, Evaluation Specialist, Appalachia Educational Laboratory
Dr. Ermel Stepp, Evaluation Specialist, Appalachia Educational Laboratory
Dr. Charles E. Kenoyer, Evaluation Specialist, Appalachia Educational Laboratory

Discussants

Dr. Gary Borich, Director of Evaluation, Research and Development Center for Teacher Education, University of Texas, and Chairman, Association of Laboratory and Center Evaluators
Dr. Desmond L. Cook, Director, Educational Program Management Center, The Ohio State University

Organizer

Charles L. Bertram, Appalachia Educational Laboratory
P. O. Box 1348, Charleston, West Virginia 25325

A Symposium of Division D, American Educational Research Association Annual Meeting, New Orleans, Louisiana

February 27, 1973

Table of Contents

Section		Author	Page
I.	Introduction	Charles L. Bertram	1
II.	A Model for the Evaluation of Educational Products	Charles L. Bertram	3
III.	The Application of an Evaluation Model to a Preschool Intervention Program	Brainard W. Hines	27
IV.	An Evaluation Model for a Regional Educational Service Agency: Construction in a Product Development Setting	Ermel Stepp	47
V.	The Application of the Evaluation Model to a Career Education Product	Charles E. Kenoyer	75
VI.	Interaction of Discussants, Panel, and Audience (Transcript)	Gary Borich, Desmond L. Cook, Panel, and Audience	91

The Application of a Model for the Evaluation of Educational Products

I. Introduction

Institutions with responsibility for developing educational products are under ever increasing pressure to "evaluate" the products which they develop. For example, the National Institute of Education is requiring the Educational Laboratories and R & D Centers to provide evaluation reports for most programs indicating that the programs should proceed through succeeding milestones (Glennan, January, 1973).

The topic of the symposium is based on the assumption that a set of evaluation activities can be stated which have applications across several programs. Another assumption on which the following objectives of the symposium are based is that professional evaluators would be in a more defensible position if there was consensus concerning the general acceptability of various evaluation procedures in given situations.

The objectives of this symposium are as follows:

1. Describe a model for the evaluation of educational products.
2. Describe experiences with using the model to evaluate selected comprehensive systems of education.
3. Consider the appropriateness of the evaluation model in light of experience gained through evaluating three educational products.
4. Arrive at alternative strategies to those described in the model for evaluation.

For this symposium, a model is considered to be an ideal sequence of activities leading to a desired end rather than an illustration depicting

perceived relationships such as a mathematical model. A set of activities leading to informed decision making will be discussed as an ideal evaluation model, and this discussion will be followed by a realistic description of the use of the model with the development of three educational products. The first product is an early childhood education program, and the model was partially a result of this effort since initiation of the early childhood program preceded the model. A second product is one designed to increase the effectiveness of school superintendents through participating in an Educational Cooperative, and third product to be discussed is a Career Education Program. Hopefully, a comparison of the ideal with the real will assist in altering the ideal model and improving the practice of product evaluation.

ED 080607

SECTION II

A MODEL FOR THE EVALUATION OF EDUCATIONAL PRODUCTS

Charles L. Bertram

TM 003 126

Appalachia Educational Laboratory, Inc.
P. O. Box 1348
Charleston, West Virginia

February 27, 1973

SECTION II

A Model for the Evaluation of Educational Products

by

Charles L. Bertram, Director of Research and Evaluation
Appalachia Educational Laboratory
Charleston, West Virginia

A statement must be made concerning the environment in which the evaluation activities are conducted before evaluation can be discussed, or even defined. The environment is educational development, and the elements of the environment include planning, product development, and diffusion as well as evaluation.

Educational Development

Educational development is the systematic process of creating and diffusing alternative products that will contribute to the improvement of educational practices (U.S.O.E., 1971). Diffusion is the process of exporting educational products, both during and following the designing and construction of the products. The result of doing educational development is educational products, which is defined as exportable methods and materials which will produce specified outcomes with designated target populations. By definition, educational development is both the creation and marketing of educational products, and to be successful, marketing must start long before the product is to be used by the target population.

A Model for Educational Development

Most institutions which do educational development have a model to guide their development activities (Barnes; Curtis, p. 42; Edmonsten; Hess and Wright; Klein and others, 1972; Research for Better Schools, Inc.; Jack Sanders; James Sanders and Worthen; Scriven and others, 1971). Most of them include phases such as needs assessment, planning or design, preliminary testing, field testing, operational testing, and usually, dissemination or the more inclusive term, diffusion.

The Model of Educational Development into which the evaluation model maps is presented as Figure 1. This seven stage model has been used by the Appalachia Educational Laboratory (AEL, 1971, p. 3). The first three stages are Needs Assessment, Feasibility Analysis, and Program Planning. These tend to be continuous, overlapping, and non-linear activities by which program plans are generated and submitted to funding agencies. Evaluators participate in the activities of these three stages since an evaluation plan is usually required to be included in the program plan.

The second set of three stages are Design and Engineering, Field Testing, and Operational Testing. The prototype product is designed and preliminarily tested under very close institutional supervision during the Design and Engineering Stage. The testing is usually of elements of the product and then combinations of elements, until the total product has met performance standards sufficiently to be field tested with a larger sample of the target population. Revisions of design during the Design and Engineering Stage have included a restatement of original goals, changes in the structure of the product or the

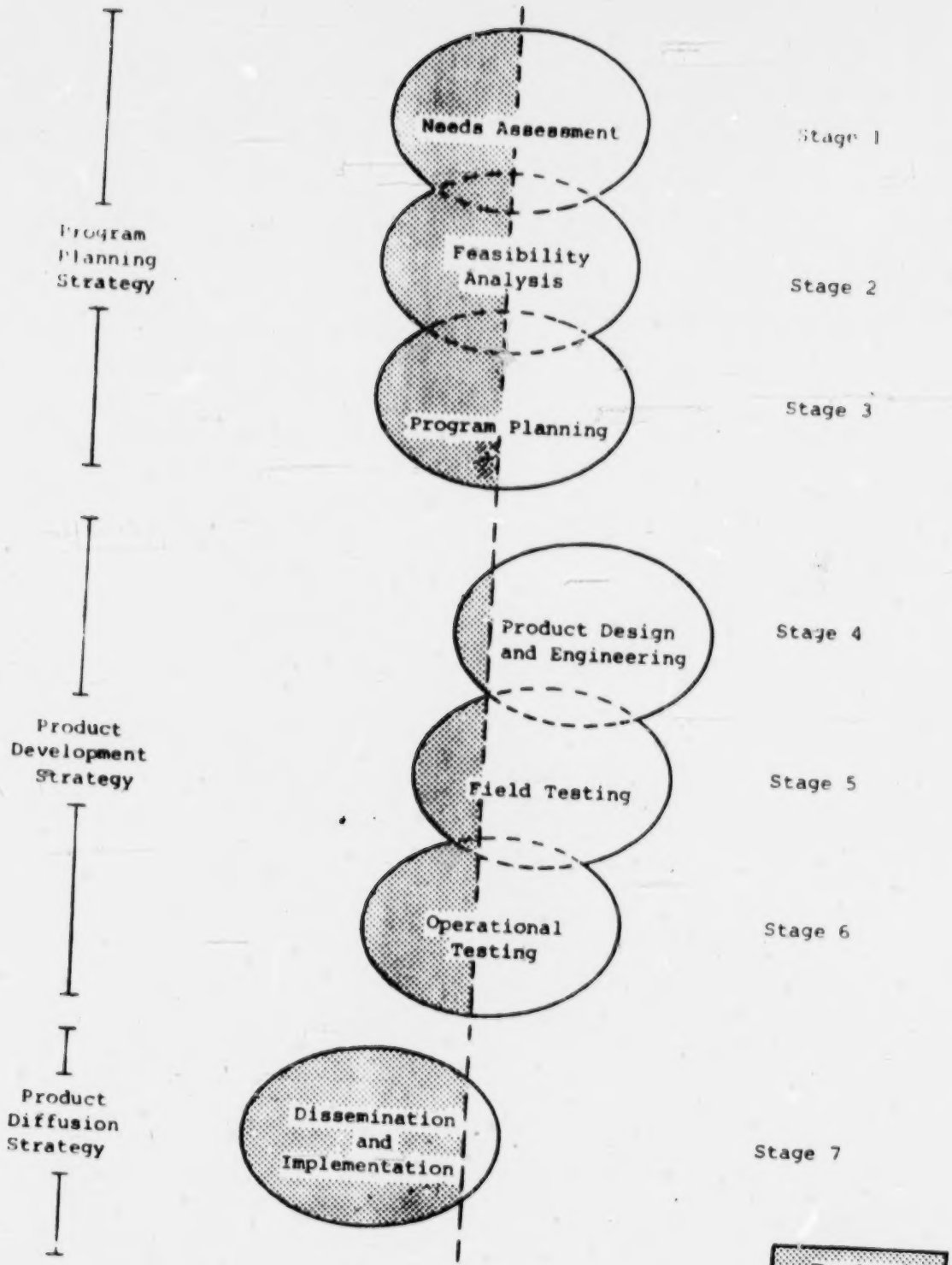


Figure 1

AEL MODEL FOR EDUCATIONAL DEVELOPMENT

process by which it is used, and changes in the evaluation procedures used to measure the degree to which the product performance goals are met.

The purpose of the Field Testing Stage is to test the product under direct control of the institution responsible for developing the product. The product is tested with a subset of the target population in a setting approximating a typical educational environment to ascertain whether the product can produce stated outcomes. The emphasis shifts somewhat from development to diffusion during Field Testing, and the results of evaluation concern both the product developer and the clientele anticipating use of the developed product. The Operational Testing Stage is to test the product with a minimum of Laboratory control and in a variety of circumstances to which the product may be adapted.

The Dissemination and Implementation Stage is to achieve widespread implementation of the product by capitalizing upon the readiness for adoption by regional constituencies. The readiness for adoption is fostered during earlier stages of development.

The duration of the educational development stages varies substantially according to the product being developed. Frequently, but certainly not ideally, one element of a product is being Field Tested while another is yet to be designed. The size of the sample during the Design and Engineering Stage is generally small. The sample size increases as the product moves into Field Testing and Operational Testing.

The preceding discussion of Educational Development has been intended as a brief description of the environment in which product evaluation exists.

Although it may sound bland, not all suggested evaluation strategies and procedures discussed in this paper will be appropriate for all developmental efforts. A suggested maxim is that "evaluation must be tailored to the product".

Rationale for the Evaluation of Educational Products

Product evaluation is the process of obtaining and providing useful information for judging decision alternatives concerning revision, disposition, and adoption of products (See Stufflebeam, p. 40). The process is based on three assumptions: (1) that product development deals with changes in product design as it is being developed, with decisions made by institutional management or funding agencies regarding the disposition of the total product development effort, and with decisions concerning adoption and implementation of the total product after it is developed; (2) that a program plan should contain information about the desired output and outcomes of product development and that changes in both product description and anticipated effects of using the product may (and should) be somewhat altered during development (Hemphill, p. 191; Scriven); and (3) that the methods at the disposal of the evaluator include such diverse tools as automatic data processing, cost analysis, opinion sampling, research design, systems analysis, statistics, and testing.

Given these three assumptions, the functions of product evaluation are to:

1. Provide to product development teams information concerning the degree to which product components are reaching intermediate objectives.

2. Provide to management of the parent institution (and funding agencies) information concerning the degree to which product components are reaching intermediate objectives and information concerning the ultimate effectiveness of the total product.

3. Provide information for product users to support implementation decisions.

An ordering of activities is implied by the evaluation functions (Figure 2) and these activities will be elaborated upon as the various activities of the evaluation model are described. In general, evaluation activities occur in the following order and recur in cycles as the product is moved through the Design and Engineering, Field Testing, and Operational Testing:

1. Assist in formulation and/or revision of product goals and objectives.

2. Select and/or revise the categories of information needed to support most critical decisions (criterion variables).

3. Select specific indicators of the performance levels associated with each criterion variable (performance indicators).

4. Describe the process by which specific data will be obtained (measurement procedures).

5. Obtain data, preferably before and after the product has been used.

6. Analyze the data and document other information obtained through the use of prototype product.

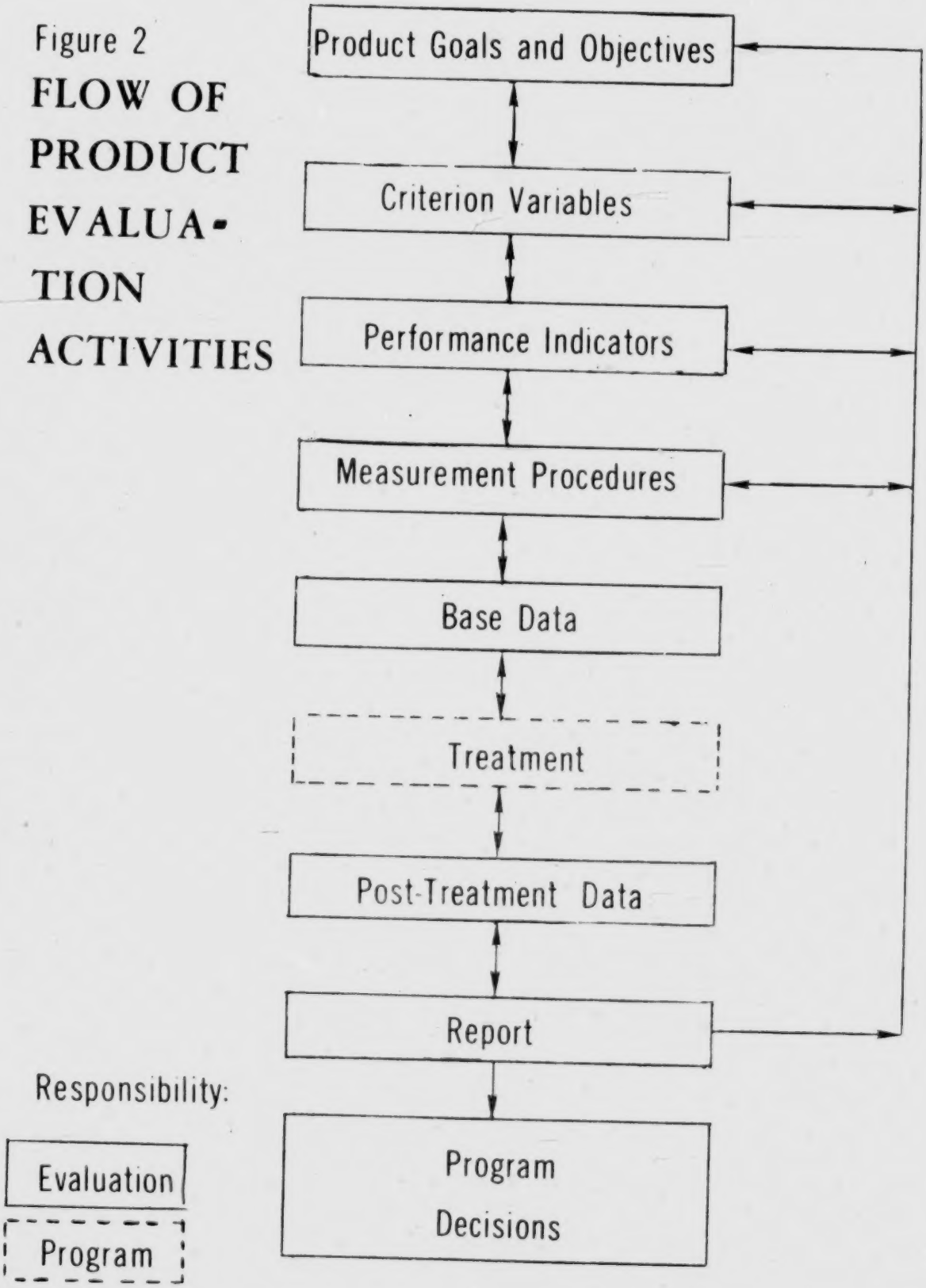
7. Organize results into the most understandable form to provide information about product effectiveness.

8. Report progress as product development is recycled or proceeds through stages of educational development.

The product evaluation model presented in the following discussion is based on experience with five product development efforts, which vary from an in-school type curricular offering in career education to a product with school administrators for a target population. In the latter case,

Figure 2

FLOW OF PRODUCT EVALUATION ACTIVITIES



an Educational Cooperative is composed of several local education agencies and a Field Test sample size of only two Cooperatives can be funded and evaluated. With the Cooperative, pre and post-testing in the conventional sense are completely inappropriate, and the "control" groups of conventional experimental design are impossible. Alternative evaluation designs must either be discovered through literature search (Glass) or invented (Stepp).

Description of the Model for Product Evaluation

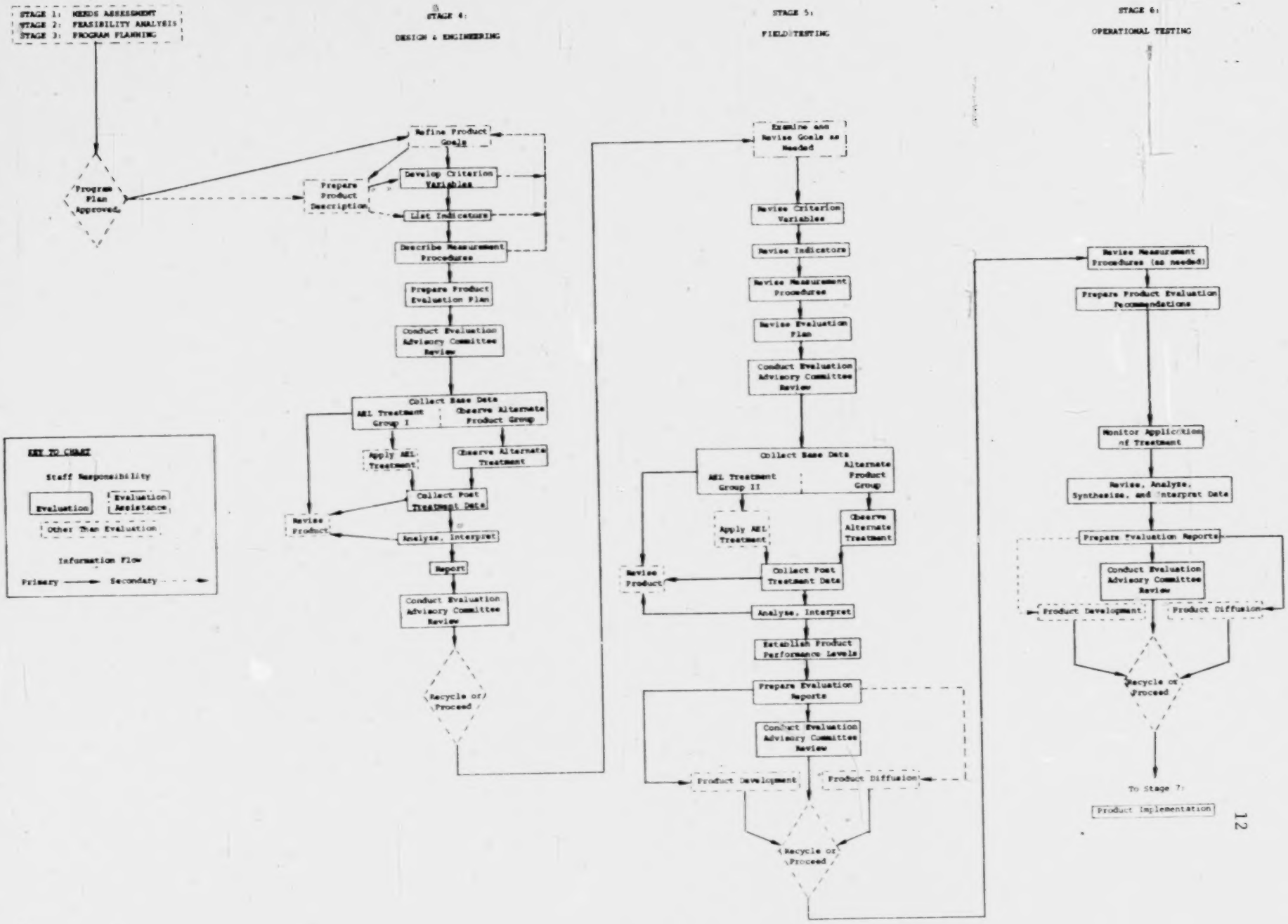
The Model for Product Evaluation as depicted in Figure 3 is a series of activities designed to produce information useful for making decisions concerning the disposition of the product. Many of the activities occur simultaneously and some activities, e.g., designing measurement procedures, are often only partially completed during one cycle and frequently revised during the next.

The educational development activities of the Needs Assessment, Feasibility Analysis, and Program Planning Stages result in a program plan which should include fairly well-defined product goals, a description of the structure of the product and how it is to be used, required funding, development time lines, plans for evaluation, and, ideally, a set of behavioral objectives based on identified educational needs. Evaluators should assist in the formulation of the program plan. This program plan, if accepted and funded, is a source document for the product evaluation staff.

The three stages of development of primary concern here are Design and Engineering, Field Testing, and Operational Testing. The development activities may recycle one or more times through a given stage before product performance will permit movement to the next stage.

AEL MODEL FOR PRODUCT EVALUATION

10 20 73



Product Evaluation During the Design and Engineering Stage

The important outputs of evaluation activities during the Design and Engineering Stage are a detailed plan for evaluation of the product, a set of measurement procedures including necessary instrumentation to determine the product's effectiveness, and information concerning the success of the prototype product in meeting objectives in a simulated environment.

Product goals. Frequently, considerable refinement of product goals is required before adequate procedures for measuring product effectiveness can be selected or designed. Goal refinement is usually achieved through face-to-face interaction of the evaluator(s) and other product development staff, and often occurs at both the total product level and the sub-product level. An indication of the relative importance of the product goals is helpful in communicating product effectiveness (Klein, 1972).

Product description. The product description should be a part of the original program plan, and serves as a guide to the evaluators in designing measurement procedures. Ideally, product goals would give sufficient guidance, but a description of the product has been found most helpful in determining if the product is being developed and used in accordance with the original design.

Criterion variables. Decision-makers usually prefer to base their decisions on a few categories of highly pertinent information. Individuals can quickly become lost in a maze of tests, subtests, treatment groups, levels, and measures occurring over several time intervals. Although the evaluation may of necessity be based on a complex design, the results are

more efficiently conveyed by a few broad categories of information which are called criterion variables. A criterion variable may be supported by several indicators. Examples of criterion variables are motor skill development, vocational maturity, or the amount of resources required to use the product.

The criterion variables are derived through consultation with product developers, by review of literature to determine what areas others have measured, and by a survey of potential consumers to determine the types of product information they desire. To be of most use, the survey of potential consumers must be completed near the beginning of the Design and Engineering Stage. The criterion variables often become section headings of evaluation reports and the subject of technical reports.

Indicators. Indicators are primary units of observation which reveal the degree to which a product has reached an acceptable level of performance. For example, an indicator of motor development in preschool children is the Marianne Frostig Test of Perceptual Development. Indicators are often based on instruments and other measurement procedures designed by the institution engaging in educational development (Bertram, Hines, & MacDonald, 1971).

Measurement procedures. The real purpose of the preceding evaluation activities is to produce efficient measurement procedures or sets of activities to obtain data for product indicators. The procedures must usually support a broad range of indicators and the range is limited only by the evaluator's (and his consultant's) imagination.

Product evaluation plan. The product evaluation plan is a statement of the results of the preceding activities, plus a schedule for completing

the evaluation, and an indication of specific staff assignments and responsibilities. The purposes of the plan are (1) to document the evaluation intent, (2) to promote a consensus among developers, diffusers, evaluators, and administrators of the educational development institution, (3) to provide a guide for completing the evaluation, and (4) to give the Evaluation Advisory Committee (mentioned below) a statement on which to base their reactions.

The plan gives the product description, including goals and objectives, in enough detail so that the plan can stand alone, and also lists criterion variables, indicators, and measurement procedures. The plan, as prepared during the Design and Engineering Stage, is revised at least annually as product revisions occur.

Evaluation Advisory Committee review. An Evaluation Advisory Committee is formed for each product being developed and meets once or twice a year. The committees are composed of from three to five members each. Selected committee members should have competencies in evaluation, research design, measurement, statistics, or the content area. Members may also represent product users, such as school superintendents, state department of education officials, and employers.

The purpose of the committees is to (1) give advice concerning more appropriate evaluation procedures, (2) serve as a stimulus to the evaluation staff, and (3) add credibility to the entire product development effort. The evaluation plan is revised, if needed, to reflect the input of the advisory committee following a thorough review.

Base data. Base data are the facts and figures which indicate the initial level of performance of the target population as measured by the

product indicators. Data are obtained from a sample of the target population which will use the prototype product and from samples who are to use alternate products, including those who will use no identifiable product at all.

Apply treatment. The application of the product and alternate treatment is monitored by the evaluators. Arranging for testing sites is a diffusion responsibility, and product development teams introduce the product to the target samples.

Collect post-treatment data. Many products move through annual cycles and, therefore, pre-treatment data can be collected in the fall and post-treatment data in the spring. Some products may require a systematic collection of data throughout the year, e.g., local board of education minutes.

Analyze, synthesize, and interpret data. Following post-treatment collection, the data are analyzed as directed by the evaluation plan. The analysis may be a simple percentage tabulation, a structuring of certain facts about the product development effort, or a rather complex analysis completed through the use of a large computer system. The results of the various analyses are organized into the most meaningful way, and the interpretation is an effort of a total evaluation team.

Problems with the treatment of data result in failure to meet production schedules more often than with any other evaluation activity. The information resulting from the interpretation of the analyses is always needed very soon after the data are collected, or the evaluation results cannot be used to suggest changes in product designs or as a basis for decisions regarding implementation of the product. The evaluator should

be especially careful to insure proper screening of data, selection of appropriate "canned" computer programs, proper labeling of subtests, and accurate recording of the age of children (such as at time of pretest rather than at different times). More problems are usually encountered in the Field Testing Stage (in which larger samples are required) than during the Design and Engineering Stage.

Prepare initial evaluation report. The primary recipient of the first evaluation report is the product development staff and the administration of the parent institution. One purpose of the report is to document a comparison of the product as originally designed with the product as used in preliminary testing. Deviations from the original design may be necessary, but they should be noted. The comparison is of both the product structure (what it is) and the process (how it is used).

A second purpose of the initial evaluation report is to document the performance of the product as it was used during preliminary testing. The changes (or lack of changes) in behavior of the target samples are recorded and comparisons with groups using alternate products are usually made. An estimate of the resources required to use the product is included, as well as the results of a study of product receptivity among potential users.

Conduct Evaluation Advisory review. The Evaluation Advisory Committee is invited to review the evaluation report while it is in draft form and to make suggestions concerning additional data analyses, different interpretations, and changes in reporting style to more efficiently communicate evaluation results. The committee members usually require one day for review of materials before interacting with evaluation and product development staff. The interaction requires approximately two days. They prepare

a written report which is communicated to administration, product development teams, and possibly to funding agencies. The report may be included as an appendix to the evaluation report. Their suggestions usually range from a general statement about the readability of the evaluation report to, as with one review, suggestions for a different post analysis of variance test.

Recycle or proceed. The decision to either proceed or return to a previous stage of the development sequence is based on several considerations, of which evaluation results may be only one factor. Recently acquired diffusion information, such as a change in potential marketability, is another factor. A third is the ability of the product development staff (as judged by Laboratory administration) to proceed with the development activities. A fourth factor regarding the continuation of the product development effort is the availability of sufficient funds or the degree to which the product conforms with priorities established by the National Institute of Education or other funding agencies.

According to the previously described Model for Educational Development, the criteria for advancement from the Design and Engineering Stage to the Field Testing Stage are (1) documented high efficiency of the product in producing specified outcomes in a limited, simulated environment and (2) evidence that the product is consistent with the potential users' needs and capabilities.

Product Evaluation During the Field Testing Stage

As indicated in Figure 3, evaluation activities encountered in the Field Testing Stage are similar to those in the Design and Engineering

Stage. The major differences between the Design and Engineering and the Field Testing Stages are that revised and validated measurement procedures are used, the product is introduced to larger samples, and formal evaluation reports document the results of the Field Testing Stage.

The goals are examined by both the evaluation and the product development teams to determine if changes, deletions, and/or additions are appropriate. The reasons for changes may be that unanticipated uses are found for the product, an unexpected market potential is found, certain former goals were not easily achieved in a cost effective manner, or the funding agency requires changes in the product. The criterion variables, indicators, and measurement procedures also need revision since revised goals imply different areas of measurement, and one purpose of the preceding stage was to produce effective measures of product performance.

The evaluation plan for Field Testing is a revision of the previous one and also includes a brief sketch of the evaluation results from the previous cycle. The cycle may have been the Design and Engineering Stage or a previous cycle of the Field Testing Stage.

The revised plan serves as a statement to which the Evaluation Advisory Committee reacts, and additional revisions are based on their recommendations. The committee should be composed of the same personnel during the different reviews if possible. The average change in personnel has been about one person per two years for Evaluation Advisory Committees.

The base data are collected on larger samples than during the previous stage, and more demands are placed on organization and data storage capabilities. For some products, major testing programs must be organized, testers trained, and provisions made for scoring tests, coding scores, and

keypunching data cards. Caution should be used to insure the compatibility for data card formats for pre and post-testing.

One purpose of collecting and analyzing data during Field Testing is to establish product performance levels. By the termination of the final cycle of Field Testing, the potential user should be informed that (1) the product will perform at a specified level on the various indicators if it is used as specified, and (2) a described amount of resources will be required to adopt and use the product.

The evaluation reports during Field Testing fulfill a dual purpose: (1) they indicate needed changes in the product and (2) they provide users with an indication of product performances as noted above. The reports have more variability in style than in the previous stage and may include technical reports, summary evaluation reports, brief descriptions of findings, and verbal presentations. The intended audience for the technical reports is research and evaluation persons in public schools, state departments of education, and institutions of higher education. The summary is intended for curriculum supervisors, teachers, and others who may not require the technical backup. A brief synopsis of evaluation results is of most use to those who need only an introduction to the results, those who have very busy schedules, and perhaps, those who make the final decisions about using products.

The Evaluation Advisory Committee review is conducted as previously described. A workable format is for an administrator to describe briefly the institution's mission and Model for Educational Development, for a representative of the development staff to describe the product or program events since the most recent review, and for the evaluation team to discuss the pertinent evaluation reports.

The decision to proceed from the Field Testing to the Operational Testing Stages is an administrative one and is again based on diffusion and evaluation considerations. The advancement criteria are (1) evidence that the product meets specifications and a high probability that it will produce specified outcomes in an Operational Test, and (2) evidence of interest in the product on the part of regional constituencies.

Product Evaluation During the Operational Testing Stage

Only minimal revisions in the product are expected during the Operational Testing Stage, and the emphasis is shifted to diffusion of the product. Supervision of product use is the responsibility of the user, and the parent institution is responsible only for monitoring the product use and effectiveness at selected demonstration sites. The evaluator's responsibility is for suggesting measures of product effectiveness, receiving and analyzing data, and reporting product effectiveness.

The first evaluation activities of the Operational Testing Stage are making any needed revisions in measurement procedures and preparing product evaluation recommendations. The purposes of preparing recommendations for the product evaluation to be conducted by the user are (1) to permit the parent institution to monitor without close supervision, (2) to give the user an opportunity to evaluate the product and thereby gain confidence in the evaluation results, and (3) to permit staff of the parent institution to determine if product performance standards continue to be met with minimum institutional control.

The recommendations are usually organized so that the user can either conduct a minimum or maximum evaluation effort, i.e., ranging from the use

of one or two instruments to a full battery of instruments and other measurement procedures, including forms for obtaining cost data. The evaluations conducted by the demonstration sites are usually much less sophisticated than those conducted during Field Testing.

The base data may need to be updated to reestablish a base against which to continue measuring growth resulting from product use. The application of the treatment is monitored through visits to demonstration sites by the evaluators. Deviations from intended product use are recorded for inclusion in the final evaluation report.

The evaluation reporting style shifts from the formal technical reporting style of Field Testing to less formal interpretive evaluation summaries. The reports must remain credible, but the audience shifts from technically-oriented to consumer-oriented groups as the diffusion function of educational development becomes more prominent than in previous stages. Reporting methods may include evaluation summaries, publications in journals, and oral presentations.

An Evaluation Advisory Committee review may be conducted following the completion of the final evaluation report. If the educational development process is successful, the product is implemented by the target population following the Operational Testing Stage.

Product Evaluation During and After Implementation

Unfortunately, the evaluation activities usually must terminate when the other product development activities have been completed, because the funding agency chooses to support another product development effort. Therefore, funding is usually not available for follow-up studies or continued monitoring of product effectiveness. One possible solution is for

the educational development institution to contract to perform the evaluation of the product at the various implementation sites. Advantages to the institution are that (1) effective procedures for implementing the product can be noted and communicated among users, (2) the institution may be made aware that needed revisions in a product require a new development effort, and (3) the evaluation talent assembled and trained for evaluation of a particular product can be used for the benefit of both the institution and the consumer. One major disadvantage with contracting to do evaluation is that attention is diverted from product development to what is essentially a service function to a broad clientele.

Summary

An outline of selected strategies and procedures used to evaluate educational products has been presented. The outline has indicated how an evaluation unit functions within an educational development organization, and has described a sequence of evaluation procedures referred to as an evaluation model.

The activities of the model are cyclical in nature, and the cycles usually require one year for completion. Major categories of activities include goal refinement, formulation and organization of measurement procedures, data collection, analysis, and reporting. One additional feature is that for each product, an Evaluation Advisory Committee examines an evaluation plan before measurements are taken, and then reacts to the evaluation reports as they are completed.

Note: An expanded version of this paper is scheduled as a chapter in a book to be published by Educational Technology Press and edited by Gary Borich.

REFERENCES

References

Appalachia Educational Laboratory, Inc. Abstract of AEL model for educational development. (Unpublished paper) Charleston, W. Va.: The Laboratory, 1971.

Barnes, E. G. The product development/diffusion process. Atlanta: Southeastern Education Laboratory, June, 1970.

Bertram, C. L., Hines, B. W., & MacDonald, R. R. Summative evaluation report: Appalachia Preschool Education Program. Charleston, W. Va.: Appalachia Educational Laboratory, 1971 (ERIC No. ED 052 837).

Curtis, W. H. Educational resources management system. Chicago: Research Corporation of the Association of School Business Officials. (Evaluation designs, models, and lists of evaluation activities are presented in Chapter VII.)

Edmonsten, L. P. A review of attempts to arrive at more suitable evaluation models: An introspective look. (Paper presented to the 1972 American Educational Research Association Annual Meeting, Chicago) Austin, Texas: Southwest Educational Development Laboratory, April, 1972.

Glass, G. V. The growth of an evaluation methodology. (Mimeographed). Boulder, Colorado: Laboratory of Educational Research, University of Colorado (undated).

Glennan, T. K. Resource allocation and management plans. (A memorandum addressed to Laboratory and Center Directors), National Institute of Education, Washington, January 12, 1972.

Hemphill, J. K. The relationship between research and evaluation studies, Ralph W. Tyler (Ed.), Educational evaluation: New roles, new means, Chicago: University of Chicago Press, 1969.

Hess, R. J. & Wright, W. J. Evaluation strategies as a function of product development stages. (Paper presented to the 1972 American Educational Research Association Annual Meeting, Chicago) St. Ann, Mo.: CEMREL, Inc., April, 1972.

Klein, S. Procedures for comparing instructional objectives. Report Number 76. Los Angeles: Center for the Study of Evaluation, UCLA Graduate School of Education, 1972.

Klein, S., Fenstermacher, G., & Atkin, M. C. The center's changing evaluation model, Evaluation comment (January, 1971), Center for the Study of Evaluation, Los Angeles.

Pena, D. Analysis of social skills development in the Appalachia Preschool Education Program. Technical Report No. 18, Charleston, W. Va.: Appalachia Educational Laboratory, 1971.

Provus, M. Evaluation or research, research or evaluation, Educational Technology, August, 1970, 50-54.

Research for Better Schools, Inc. Educational development processes (Paper presented to the 1970 Annual Directors' Conference, San Diego, California.) Philadelphia: Research for Better Schools, Inc., December 8, 1970.

Sanders, J. & others. Instructional development at the Ohio State University Evaluation Center (draft), Columbus: The Ohio State University Evaluation Center, August, 1972.

Sanders, J. R. & Worthen, B. R. A descriptive summary of frameworks for planning evaluation studies, SRIS Quarterly, Bloomington, Indiana: Phi Delta Kappa, Spring, 1972, 10-14.

Scriven, M. Presentation to Conference of Educational Development and Research (CEDaR) directors' meeting in Phoenix, Arizona, January 18, 1972, during which "goal free" evaluation was discussed.

Scriven, M. Pros and cons about goal-free evaluation, Evaluation comment, Center for the Study of Evaluation, Los Angeles, December, 1972, 1-4.

Scriven, M., Glass, G. V., Hiveley, W., & Stake, R. E. An evaluation system for regional labs and R and D centers. Washington: U.S.O.E. August 31, 1971. (ERIC No. ED 061 299).

Stepp, E., Jr. General evaluation design for the Educational Cooperative. (draft, mimeographed). Charleston, W. Va.: Appalachia Educational Laboratory, 1972.

Stufflebeam, D. L. Educational evaluation and decision-making. Itasca, Illinois: F. E. Peacock Publisher, 1971.

U. S. Office of Education, National Center for Research and Development. Information and instructions for submittal of the Basic Program Plan for educational laboratories and research and development centers. (Mimeographed). Washington: U.S.O.E., June 3, 1971.

ED 080608

SECTION III

THE APPLICATION OF AN EVALUATION MODEL TO A PRESCHOOL INTERVENTION PROGRAM

Brainard W. Hines

TM 003 127

Appalachia Educational Laboratory, Inc.
P. O. Box 1348
Charleston, West Virginia

February 27, 1973

SECTION III

The Application of an Evaluation Model to a Preschool Intervention Program

The purpose of this paper is to relate the evolution and application of a model for evaluation to a preschool intervention program. Essentially it will be oriented to the historical and factual level, rather than to the more abstract discussion of the evaluation model. It will attempt to give an historical overview of Appalachia Educational Laboratory's preschool intervention program as it has moved through the development process and will concentrate on the effect of evaluation on that process.

The paper, then will have two major goals. It will attempt to point out congruencies between product development and evaluation methodology, and will also provide a basis for planning further evaluation activities in the area of preschool programs both at AEL and in the larger educational community.

Program Description

The Home-Oriented Preschool Education (HOPE) program, as developed by AEL, is a multi-component approach to Early Childhood Education. As originally conceived, the program was intended as an alternative to traditional kindergarten programs, which were not practicable for use in rural Appalachia. It was the Laboratory's intent to use technology in the form of television programming, a mobile facility, and paraprofessional home-visitors to provide a broad range of learning experiences for the children in Appalachia.

However, as the program developed, its general goals expanded beyond the area of an alternative to public kindergarten. Although this is still an important area of concentration, it was anticipated that the HOPE process

would provide for the needs of preschoolers which were not currently being met by traditional kindergartens.

The three parts of the HOPE process each address themself to a specific target area in the child's environment or his behavioral repertoire. The purpose of the Home Visitor is to not only work with the child himself, but also to make lasting changes in the parent-child relationship which will maintain the child's achievements. The overall curriculum of the television program stresses not only cognitive and perceptual motor skills, but also affective learning and social development. The mobile facility stresses many of these objectives and emphasizes development of the social skills necessary for cooperative group behavior.

The target population for the HOPE Program consists of those 3-, 4-, and 5-year-old children within the Appalachian region who do not have access to quality educational experiences. Although the target population was not specifically limited to the highly related and rural Appalachian child, it was intended to reach that group as fully as possible.

Early History and Development of the HOPE Program

The development of the idea for a Home Oriented Preschool Education Program was an evolution from general regional needs to specific solutions for a specific population. After the first long-term funding of AEL in June of 1966, a number of regional needs and priorities were identified. One of the most pressing of these needs was the necessity for overcoming the effects of regional isolation on the children of Appalachia.

The nationwide interest in Early Childhood Education led the AEL staff to examine the effectiveness of existing programs, such as Headstart and more traditional kindergartens in meeting the needs of preschool children

in Appalachia. The results of this study indicated that poor roads, insufficient teacher availability and the lack of available funds argued against more traditional types of preschool programs.

The idea of using a home-oriented, three-prong approach to early childhood education was fully accepted by AEL in 1967 as a proposed program start. In order to derive a suitable preschool curriculum, a subcontract was established between AEL and the College of Human Resources and Education of West Virginia University. This contract consisted of the performance of an extensive review of the existing literature on preschool education, a series of behaviorally oriented objectives in a broad curricular range, and a study of the needs of preschool children in Appalachia. Although it was originally intended to turn the entire production of the ECE program over to the staff of Human Resources Department, financial and logistical considerations prevented this from taking place.

After the decision was made to produce and evaluate the ECE program directly under Laboratory supervision, it was necessary to locate, hire, and train the staff needed to transform the behavioral objectives of the curriculum into specific teaching activities.

At this point it will be well to examine the previously related history of the program to the current evaluation model. The needs assessment had been done by a consensus of individuals knowledgeable in the educational problems of the region, and a goal had to be agreed upon in the area of early childhood education. No empirical data were gathered on the general need for such a program, and although a preliminary decision had been made to utilize a home-oriented approach, no empirical support was present for the acceptability of such a program. The necessity of getting a program started

and in operation as soon as possible made the empirical investigation of such questions impractical.

Also, a decision was made that the West Virginia University environment would provide a richer source of personnel, production facilities and materials than would be available at the central Laboratory location. In this sense a feasibility study was conducted, and it was decided that the Laboratory could not undertake a venture of this scope on such short notice. However, when final production schedules and budgetary proposals were received from West Virginia University, it was apparent that not only would this alternative be too expensive, but more important, the production schedule would take far longer than was acceptable.

Therefore, despite the difficulties involved, the Laboratory undertook the program development process itself. Not only did this necessitate hiring production and field test staff, but also the establishment of a staff for evaluating the effectiveness of the overall effort.

At this point (summer, 1968) the following segments of the program development sequence had occured: A needs and feasibility study was done. A program description had been elaborated, and a curriculum formation and preliminary definition of overall program goals were formulated. No specific evaluation plan had been formulated at that time, although the author of the curriculum preparation expressed the feeling of AEL management and development staff when he noted, "to a larger extent, the potential success of the overall (ECE) demonstration program depends heavily upon the virtues and faults of the evaluation design" (Hooper, p. 3).

The evaluation design itself was formulated to cover the three-year developmental span of the ECE program. During the program planning phase, Dr. Gerald Lesser and Dr. Warren Seibert of the U.S.O.E. site review team

stressed the overall importance of evaluation and offered general suggestions for the evaluation procedure. In order to implement these suggestions, a panel of research and evaluation experts met with the AEL staff to formulate such a plan.

It was from this meeting that the development of criterion variables, indicators of those variables, and finalization of measurement procedures took place.

The final evaluation plan (as of June, 1968) specified the following:

A. Criterion variables which were selected were in the areas of program performance (child's behavior), program effort and required resources, program performance pervasiveness (logistics of widespread adoption), and program cost analysis.

Within the broad categories other subdivisions of criteria were organized. The basic areas of language skills, psychomotor performance, verbal attainment (I.Q.), and social skills were identified as areas of child behavior to be assessed. Cost figures, resource availability, attitude of the overall population were included within the other areas of measurement.

B. Indicators of performance were selected for the above areas on the basis of availability, applicability to criterion variables, and previous utility. For the area of program performance, several nationally normed tests were chosen. These tests included the Illinois Test of Psycholinguistic Ability, The Marianne Frostig Developmental Test of Visual Perception, and the Peabody Picture Vocabulary Test. In addition, AEL constructed tests of social development and the attainment of overall program objectives were planned. Other indicators included parental surveys, questionnaires, and demographic information.

C. Measurement procedures were specified. The field test site which was selected was located in south-central West Virginia. Within that region, it was decided to locate three "treatment" groups, each consisting of approximately 150 children. One of these groups was to view the television program and receive all related printed materials, the second was to be visited by the paraprofessional and view the television program, and the third was to visit the mobile classroom in addition to the other components. Children of all three ages and both sexes were to be represented equally within each group. A measurement sample of 96 children was selected from this larger group to expedite testing.

The first major problem in implementing the evaluation plan arose with the allocation of individual families to treatment groups. Random sampling across such a large geographical area (six counties) was not possible because of the travel involved for the paraprofessionals and more particularly the problems associated with group meetings on the mobile classroom. In a rural and sparsely populated region such as this, random sampling of children was felt to be impractical.

As an alternative, a random sampling of geographic grid areas was performed, with most of the children in a given area being allocated to one of the three treatments. As was later found, however, this produced some systematic socio-economic bias into the samples, with the "TV-only" group having a slightly lower s.e.s. and being more rural than the remaining two groups.

A second major problem arose with selection of individuals to conduct the actual testing of the children concerned. At first it was felt that the paraprofessionals were best suited for this task and they were trained by the AEL Evaluation staff to administer the PPVT, Frostig, and ITPA as a pretest measure in September, 1968. As the testing progressed, it became

evident that not only did the home visitors resent this task, in many cases they were allowing their previous knowledge of the child to bias the test results to some degree.

Simultaneously with the testing in the field site, a control group in southern Virginia was located and was tested by individuals living in that area. Again, because the testers were not experienced with the instruments used, a question of validity of results arose.

Despite these problems, base-line data on the children were collected in the fall of 1968. At this time, however, it was not possible to administer a curriculum specific measure to evaluate the basic level of attainment of program objectives. This was because the final selection of the first year's objectives had not been completed, thus preventing a sampling of items. Social skill attainment was not measured because of the lack of AEL staff and resources to measure this area.

After an interval of nine months, posttest information was collected to provide an estimate of the gains produced during the first year's operation. During this interval, the first year's objectives were compiled into the Appalachian Preschool Test (APT), a curriculum specific measure. However, during this time the mobile classroom had been available for only four months, due to logistical problems in constructing the teaching enclosure.

After the first year's operation, an evaluation report was prepared, comparing the gains of the three treatment groups with those of the control group. Generally, the results indicated slight gains in most areas of measurement and equivocal changes in those areas where gains were noted.

Little positive information was provided in the first year's evaluation report on the area of program performance. Only one of the language subtests

showed significantly different gains across groups, the PPVT showed no significantly different gains, and although the three "treatments" scored significantly above the control group on curriculum specific measure, no differences were noted across the three treatment groups.

Because of the lack of definite information as to the validity of the information provided by the control group testing, and the problems associated with the use of home visitors as testers, no decision was made to recycle the entire program. Instead, efforts were made to improve the quality of the evaluation process in respect to the information needed by the management of the Laboratory.

A review of the first year's report was conducted by the panel which had first met to decide on the preliminary evaluation plan. At that time (December, 1969) many changes in the format of the report, the nature of the data analysis to be done, and in the mechanics of testing were recommended.

These recommendations resulted in the following changes in evaluation procedure.

A. Home visitors were trained and supervised by AEL staff to conduct all structured testing of children in the sample, but did not test children whom they knew.

B. The location of a new control group, closer to the Laboratory which could be more closely monitored and tested.

C. An emphasis on actual mean differences across groups instead of reliance on gain scores in future reports.

D. Simplified reporting of evaluation findings.

E. Increased evaluation staff size and diversity.

F. An emphasis on social skills measurement.

G. An increased emphasis on formative or within-program evaluation.

H. A reduced testing battery to simplify the testers' task.

It should be noted that at this point, no differentiation had been made between the design and engineering stage of development and the field test stage. Essentially, the difficulties which arose during the evaluation of the first year's operation could have been to a large degree minimized if small scale (trial run) procedures had been tested, before a more ambitious evaluation and implementation had been undertaken. To a large degree, this conclusion resulted in the structure of the first part of the current evaluation model.

Later Program Development

During the second and third year of the ECE program's development, a major evaluation event took place at the conclusion of each program year.

In June of 1970 (end of the second year) all of the children in the original sample were tested with the reduced battery of instruments. This battery consisted of one subtest of the ITPA, two subtests of the Frostig, the PPVT, and the curriculum specific measure. Additionally, a social skills measurement instrument was devised, based on an interaction analysis technique. One group of children who had visited the mobile classroom, and one which had not had that experience were videotaped in a group activity. Their behavior was subsequently analyzed for any systematic group differences.

After a preliminary report was made on findings of this testing, the evaluation staff and the advisory committee met and reversed several previous decisions. First, it was felt that the continuing problems with the use of paraprofessionals as testers necessitated the employment of individuals specifically for that function. Second, although the reduced test battery was more specifically oriented to the goals of the program, it did

not provide for the possibility of secondary (unplanned) effects. Additionally, data from the first year of evaluation had included these subtests, and if longitudinal information was needed, these subtests would have to be administered. Finally, the second control group was not adequately selected, producing mean scores which were as much as one standard deviation above the norms in some areas. This necessitated the location and testing of another group of children for the purposes of comparative analysis.

Another consideration arose because of the small sample size in each age by treatment by sex cell classification. With an original sample this provided for only six children in each classification cell. In several cases, by the end of the second year some cells were completely empty due to withdrawal of students from the program.

It was decided to administer the entire test battery to half of the children in the program in September of 1970, and to the total group in June of 1971. Additionally, a similar procedure (Solomon 4 group design) was used to test the new control group so an indication of the effect of testing could be obtained.

Location and testing of the new control group was done by West Virginia University in close contact with the AEL evaluation staff. This was to be the final segment of the program evaluation before a decision was reached on whether to proceed with operational testing, and every effort was expended to produce a meaningful and valid evaluation of the overall effects of the ECE program.

Because of the necessity for collecting further data, the actual writing of the second year's final evaluation report was delayed for almost six months. Although it followed the general format of the first report, it was judged

by the evaluation review panel as being considerably more sophisticated in both analysis and quality of language.

Where the first summative evaluation report had been written almost exclusively by one individual, the second was presented in the form of a number of individually written technical reports which could be read separately or together as the readers' needs dictated. A summary of the overall findings was also prepared to give an overview of the strengths and weaknesses of the program in its current stage of development.

The final evaluation procedure attempted to answer questions in the same areas as the original (first year) documentation of evaluation. It also addressed itself to the overall considerations of a comparison with a traditional kindergarten, one of the original goals of the program. Although limited resources were available for this area, a comparison was made on the curriculum specific measure and Peabody I.Q. test.

The final evaluation report incorporated the data on all the previously mentioned indicators of child performance, program cost, and parental attitude, and stressed a format which could meet the needs of a broad potential audience.

The final evaluation on the field test stage was completed in June of 1971, and after information had been organized and subjected to analysis, a decision was reached that the program was producing levels of performance sufficient to warrant its transition to the operational test stage.

At this point, it would be appropriate to consider the problems which arose in the application of the evaluation procedures to the field test stage of development. The major problems were as follow:

A. A comprehensive, action-oriented model had not been elaborated or made explicit enough to guide all the activities of the evaluation staff.

B. Relations between evaluation and production staff were strained at times because of differing philosophies, overlapping role expectations, and the threat inherent in an "outside" evaluation.

C. Data analysis was difficult because of the lack of physical facilities, professional resources, and the volume of materials to be scored, coded, and subjected to statistical analysis.

D. Most important of all, there often was not enough time available to gather the necessary data before a program decision had to be reached by the AEL management. The information flow in the evaluation procedure could be compared to the motion of a liquid through a pipe at fixed diameter. The transfer and process time is proportional to the volume to be moved, and to the amount of transformation necessary in the system before a decision can be reached.

E. The incorporation of a "formative evaluation" unit within the program to provide rapid feedback in a variety of areas was partially successful, but no structured record of these events was kept for future reference.

Operational Test Phase

After the decision had been made to move the ECE program (renamed HOPE at this time) to the operational test phase, several sites in the region were selected for the purpose of ascertaining the effectiveness of the program with minimal Laboratory supervision.

Evaluation procedures were suggested for these sites, but the specific adoption of given instruments, etc., was left to the discretion of the site. Evaluation staff was made available to the test sites, but no demands were made on the amount of supervision which those individuals could exercise over the program evaluation.

Basically the operational test evaluation used the same instruments which had been incorporated previously, but allowed the individual sites to select the battery and staff time for testing allowed. Funds were not available for hiring testers at the sites, so the paraprofessionals were required to test the children enrolled in the programs.

Data were gathered on a pre-post basis at all sites, and test scoring and coding were done at the site. Data analysis, interpretation, and report writing were done by the evaluation staff at AEL. This division of labor saved time, but sacrificed control and supervision of data gathering, and made enforcement of production timetables almost impossible.

The Evaluation Department also made an effort to describe any differences in program implementation across the sites, and to relate those differences to any systematic biases in the children's behaviors as measured by the test battery.

An evaluation report was prepared on all of the test sites, comparing their performance to that of the previous control group and to the performance of the children in the previous field test. It was found that the operational sites were performing at a level of efficiency approximately equal to that of the previous year, but were spending relatively more per child than had been necessary in the field test. This may have reflected the rather small numbers of children served in each site.

Discussion

Up to this point, this paper has focused on a historical documentation of actual events, and has not attempted to report on specific results or findings which were obtained from the evaluation itself. Neither has it specified the objectives, goals or prepared outcomes of the program itself.

This information is available from the documents listed in the bibliography which follows, and is not necessary for understanding the process which is involved in evaluating a given program which is undergoing development or diffusion.

Generally, it is possible to point to the strengths and weaknesses of the evaluation process which has been documented here. First, the process is designed to provide information to decision-makers in a variety of positions and thus is flexible in its format of reporting findings. The wide range of individuals in the target audience necessitates a corresponding range of report formats and writing skills from evaluation staff.

Second, the sheer volume of information required to make multiple decisions on a variety of areas and the necessary constant and close supervision of every phase of information gathering requires a large and diverse evaluation staff. With such a large staff to meet periods of heavy activity, individuals sometimes find themselves with little work as a given product moves to a stage which does not require their expertise. Again, as programs tend to follow cyclical development or evaluation activities, the work load is also seasonal in its volume.

Third, the necessity of rapid decisions in matters of great importance to the program occasionally forces circumvention of the evaluation process. That is, decisions must sometimes be made without benefit of the empirical evidence which would be available at a later time. The evaluation process and model which has been described allocated the time of a decision to the availability of information, while in reality this match is not always made.

Fourth, the constraints of funding may not permit the actual recycling of a program, process, or component which has not succeeded in meeting its goals. Changes may be suggested, but may not be implemented simply because sufficient funds and/or time are not available.

Fifth, the evaluation process in a multi-component program may not be able to specify direct relations of cause and effect between target population performance and individual segments of a program. This is particularly true in the HOPE program here described, where the three program components had overlapping goals. Thus, it is difficult to allocate responsibility for charge to a given component.

Finally, a great deal of difficulty is encountered in setting goals and objectives for a developing program and in adhering to those goals as new staff members are added to the development team, and as subjective impressions are used to determine the feasibility of the task being undertaken. As the program develops, the goals may change direction without the knowledge of the evaluation staff.

For this reason, close cooperation between development and evaluation staff is necessary not only to successful completion of the entire development process, but to the overall accuracy of the information which the evaluation was intended to deliver.

Bibliography

The historical documentation which is contained in this paper is derived from numerous letters, memos, and personal experiences of the author. A broad account of program development, evaluation procedures, and findings may be found in the following Laboratory documents.

List of Available Publications Pertaining to the Home-Oriented Preschool Education Program

Available Evaluation Reports

School Year 1968-69

1. Evaluation Report: Early Childhood Education Program, 1969 Field Test, Appalachia Educational Laboratory, Inc., Charleston, West Virginia, March, 1970, ED 041 626.

School Year 1969-70

2. Evaluation Report: Early Childhood Education Program, 1969-70 Field Test, Appalachia Educational Laboratory, Inc., Charleston, West Virginia, May, 1971. (Authors: Charles Bertram, Deagelia Pena, Brainard Hines)

3. Evaluation Report: Early Childhood Education Program, 1969-70 Field Test. Summary Report, Appalachia Educational Laboratory, Inc., Charleston, West Virginia, May, 1971, ED 052 837.

4. Demographic and Socio-Economic Data of the Beckley, West Virginia Area and 1968-1970 Developmental Costs of the Early Childhood Education Field Study. Technical Report No. 1, Appalachia Educational Laboratory, Inc., Charleston, West Virginia, February, 1971, ED 052 832. (Author: Charles Bertram)

5. Analysis of Intelligence Scores. Technical Report No. 2, Appalachia Educational Laboratory, Inc., Charleston, West Virginia, February, 1971, ED 052 838. (Author: Brainard Hines)

6. Attainment of Cognitive Objectives. Technical Report No. 3, Appalachia Educational Laboratory, Inc., Charleston, West Virginia, February, 1971, ED 052 833. (Author: Brainard Hines)

7. Detailed Analysis of Language Development of Preschool Children in ECE Program. Technical Report No. 4, Appalachia Educational Laboratory, Inc., Charleston, West Virginia, ED 052 834. (Author: Brainard Hines)

8. Analysis of Visual Perception of Children in the Early Childhood Education Program (Results of the Marianne Frostig Developmental Test of Visual Perception). Technical Report No. 5, Appalachia Educational Laboratory, Inc., Charleston, West Virginia, February, 1971, ED 052 839. (Author: Brainard Hines)

9. Factor Analysis of the Early Childhood Education Test Data. Technical Report No. 6, Appalachia Educational Laboratory, Inc., Charleston, West Virginia, February, 1971, ED 052 840. (Author: Deagelia Pena)

10. Social Skills Development in the Early Childhood Education Project. Technical Report No. 7, Appalachia Educational Laboratory, Inc., Charleston, West Virginia, February, 1971, ED 052 835. (Authors: Deagelia Pena and George Miller)

11. Results of Parent and Student Reaction Questionnaire. Technical Report No. 8, Appalachia Educational Laboratory, Inc., Charleston, West Virginia, January, 1971, ED 052 836. (Author: Brainard Hines)

12. Analysis of Children's Reactions to AEL's Preschool Television Program. Technical Report No. 9, Appalachia Educational Laboratory, Inc., Charleston, West Virginia, December, 1970, ED 052 841. (Author: George Miller)

13. A Comparison of Parents' Attitudes Toward AEL's Around the Bend and Other Children's Television Programs. Technical Report No. 10, Appalachia Educational Laboratory, Inc., Charleston, West Virginia, December, 1970, ED 052 842. (Author: Charles Bertram)

School Year 1970-71

14. Summative Evaluation of the Appalachia Preschool Education Program, Appalachia Educational Laboratory, Inc., Charleston, West Virginia, December, 1971, ED 062 024. (Authors: Charles Bertram, Brainard Hines, Randolph MacDonald, Ermel Stepp)

15. Demographic and Socioeconomic Data of the Beckley, West Virginia Area and 1968-1971 Development Costs of AEL's Preschool Education Field Study. Technical Report No. 11, Appalachia Educational Laboratory, Inc., Charleston, West Virginia, December, 1971. (Authors: Charles Bertram and Randolph MacDonald)

16. Analysis of Intelligence Scores. Technical Report No. 13, Appalachia Educational Laboratory, Inc., Charleston, West Virginia, December, 1971, ED 062 016. (Author: Randolph MacDonald)

17. Attainment of Cognitive Objectives. Technical Report No. 14, Appalachia Educational Laboratory, Inc., Charleston, West Virginia, December, 1971, ED 062 017. (Author: Brainard Hines)

18. Detailed Analysis of the Language Development of Children in AEL's Preschool Education Program. Technical Report No. 15, Appalachia Educational Laboratory, Inc., Charleston, West Virginia, January, 1972, ED 062 018. (Author: Brainard Hines)

19. Analysis of Visual Perception of Children in the Appalachia Preschool Education Program. Technical Report No. 16, Appalachia Educational Laboratory, Inc., Charleston, West Virginia, December, 1971, ED 062 019. (Author: Brainard Hines)

20. Factor Analysis of the Appalachia Preschool Education Program Test Data. Technical Report No. 17, Appalachia Educational Laboratory, Inc., Charleston, West Virginia, December, 1971, ED 062 020. (Author: Randolph MacDonald)

21. Analysis of Social Skills Development in the Appalachia Preschool Education Program. Technical Report No. 18, Appalachia Educational Laboratory, Inc., Charleston, West Virginia, January, 1972, ED 062 021. (Author: Deagelia Pena)

22. A Comparison of Parents' Attitudes Toward AEL's Around the Bend and Other Children's Television Programs. Technical Report No. 21, Appalachia Educational Laboratory, Inc., Charleston, West Virginia, December, 1971. (Authors: Charles Bertram and Randolph MacDonald)

23. Measuring Children's Curiosity. Technical Report No. 22, Appalachia Educational Laboratory, Inc., Charleston, West Virginia, December, 1971, ED 062 022. (Author: George Miller)

24. A Comparison of AEL's Preschool Education Program with Standard Kindergarten Programs. Technical Report No. 23, Appalachia Educational Laboratory, Inc., Charleston, West Virginia, January, 1972, ED 062 023. (Author: Charles Bertram)

School Year 1971-72

25. Summative Evaluation of the Home-Oriented Preschool Education Program During the 1971-72 Demonstration Phase, Appalachia Educational Laboratory, Inc., Charleston, West Virginia, December, 1972.

Home-Oriented Preschool Education Manuals

1. Program Overview and Requirements, Appalachia Educational Laboratory, Inc., Charleston, West Virginia, October, 1972. (Author: Roy Alford)

2. Field Director's Manual, Appalachia Educational Laboratory, Inc., Charleston, West Virginia, October, 1972. (Authors: George Miller and Roy Alford)

3. Handbook for Mobile Classroom Teachers and Aides, Appalachia Educational Laboratory, Inc., Charleston, West Virginia, October, 1972. (Author: Mabel Little)

4. Home Visitor's Handbook, Appalachia Educational Laboratory, Inc., Charleston, West Virginia, October, 1972. (Author: Nita Nardo)

5. Curriculum Planning Guide, Appalachia Educational Laboratory, Inc., Charleston, West Virginia, October, 1972. (Author: Marie Snider)

6. Personnel Training Guide, Appalachia Educational Laboratory, Inc., Charleston, West Virginia, October, 1972. (Author: Nita Nardo)

7. Materials Preparation Guide, Appalachia Educational Laboratory, Inc., Charleston, West Virginia, October, 1972. (Author: Marie Snider)

Other Available Publications

1. Alford, Roy, The Appalachia Preschool Education Program: A Home-Oriented Approach, December 1, 1970, ED 052 865.

2. Alford, Roy, Teaching Mathematical Concepts to Rural Preschool Children Through a Home-Oriented Program, Doctoral Dissertation, Appalachia Educational Laboratory, Inc., Charleston, West Virginia, June, 1970.

3. Appalachia Educational Laboratory: Early Childhood Education, Career Decision-Making, Educational Cooperatives, Appalachia Educational Laboratory, Inc., Charleston, West Virginia, November, 1971. (Brochure)

4. Development of the Early Childhood Education Program, Basic Plan, Appalachia Educational Laboratory, Inc., Charleston, West Virginia.

5. Marshall, H. William and Frank H. Hooper, The Initial Phase of a Preschool Curriculum Development Project, (Final Report) Appalachia Educational Laboratory, Inc., Charleston, West Virginia, August, 1968, ED 027 071.

6. Preschool in Appalachia - School Without a Schoolroom, Appalachia Educational Laboratory, Inc., Charleston, West Virginia.

ED 080609

TM 002 128

SECTION IV

AN EVALUATION MODEL FOR A REGIONAL EDUCATIONAL SERVICE AGENCY: CONSTRUCTION IN A PRODUCT DEVELOPMENT SETTING

Ermel Stepp, Jr., Ed.D.
Design and Evaluation Specialist

Appalachia Educational Laboratory, Inc.
P. O. Box 1348
Charleston, West Virginia

February 27, 1973

SECTION IV

An Evaluation Model for a Regional Educational Service Agency: Construction in a Product Development Setting

The domain of educational development as envisioned by the Appalachia Educational Laboratory, Inc., is subsetted by seven stages. The stages are grouped into three strategy groups and characterized by emphases on product development and product diffusion (Figure 1). The first group of stages in the domain is that of program planning strategy which encompasses the stages of needs assessment, feasibility analysis, and program planning. The product development strategy is the next group of three stages which includes product design and engineering, field testing, and operational testing. The seventh stage is that of product diffusion strategy and attends to dissemination and implementation.

Program Planning Strategy

Program planning strategy attends to needs assessment, feasibility analysis, and program planning. An attempt will be made to recapture the development of program planning strategy for the Educational Cooperative. That is, an examination will be made of the needs assessment, feasibility analysis, and program planning which constituted the program planning strategy for the Educational Cooperative development program of the Appalachia Educational Laboratory.

Needs Assessment

A rudimentary consensual needs assessment of the Appalachian Region was accomplished by the steering committee which proposed the establishment of the Appalachia Regional Laboratory (Steering Committee, 1965, pp. 10-11).

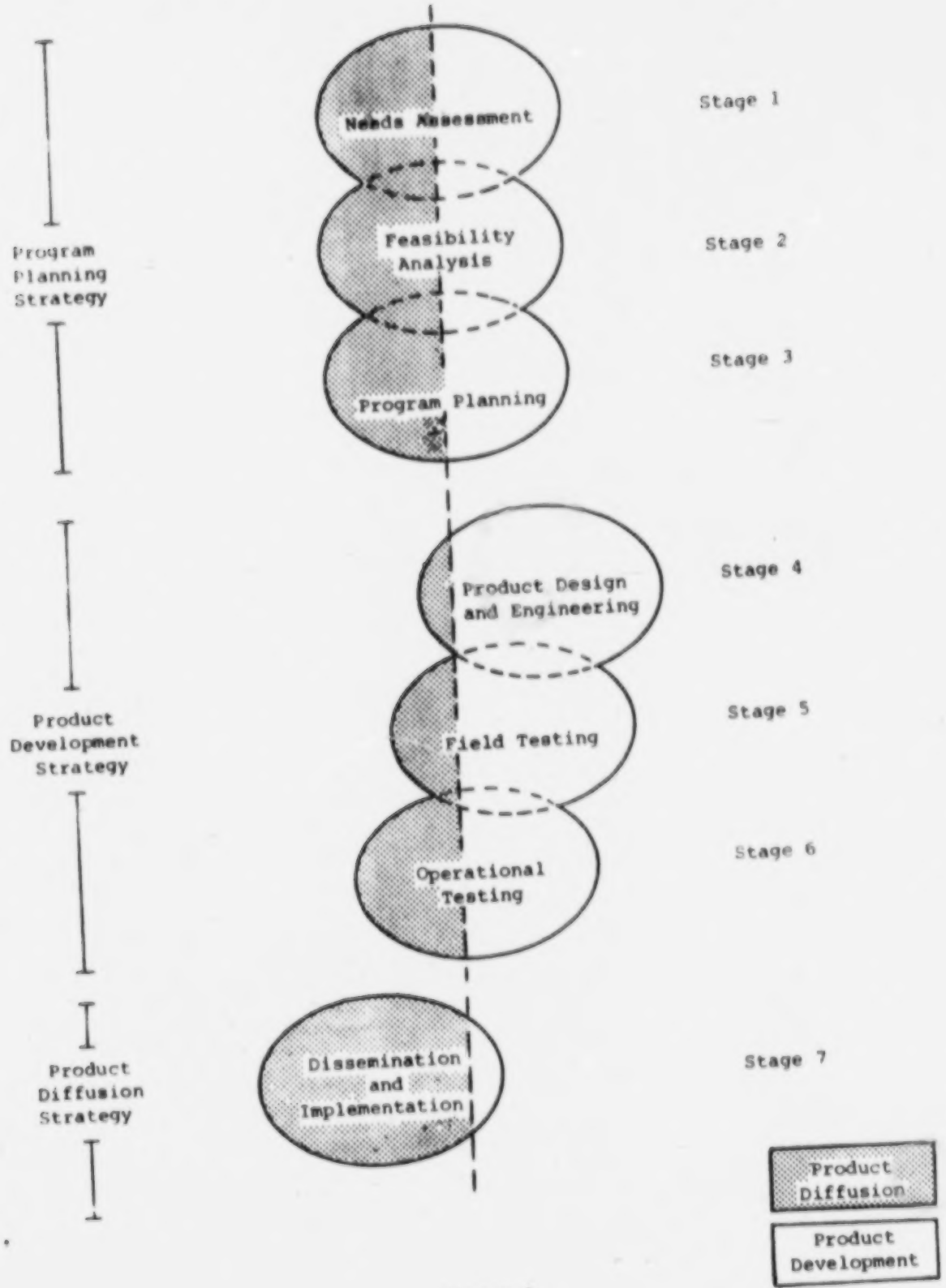


Figure 1

AEL MODEL FOR EDUCATIONAL DEVELOPMENT

> "This, then, is Appalachia and these are its educational needs. While including a significant portion of the land area and population of the United States it is a region apart from the rest. A non-urban land, also non-farm, much of it deprived of the facilities, services, and economic characteristics of modern America society. Its distinctive educational problems stem from basic deficiencies in the physical, economic, and social environment. The educational systems of the region are beset by problems of poverty and consequent cultural deprivation, outmoded curriculum materials and concepts, educational and social isolation, inadequate provisions for relating school to work, low levels of aspirations and expectations, and a resistance to educational change."

In order to have an impact on the Appalachian region (Figure 2), it was thought that a need existed for a different type of delivery system for educational innovations. The vehicle for greatest change in Appalachia appeared to be an organizational design for a multi-district cooperative system. This cooperative would become a system for the facilitation of needs satisfaction in the Appalachian region. The setting would be for great change in a low information field. Stufflebeam and others later referred to this as a neomobilistic decision setting.

The principal axis of the organizational identity to be fostered in the educational cooperative was the character of a social system with a mission as an institutional educational change agent. The educational cooperative was to be a macrosystem with an adaptive dynamism for the facilitation of school system renewal. The macrosystem was to be regional collaboration on the enhancement of systems decidability. Collaboration on systems decidability was to evoke a functional compatibility of value orientations, needs structures, resource mix, and adaptive rationality. Adaptive systems rationality was to be an operation under positive sanction of accountability and effectiveness. The facilitation of school system renewal was to involve the educational cooperative as a model of administrative practices in its systems rationality such that constituents would reallocate resources reflecting renewal.

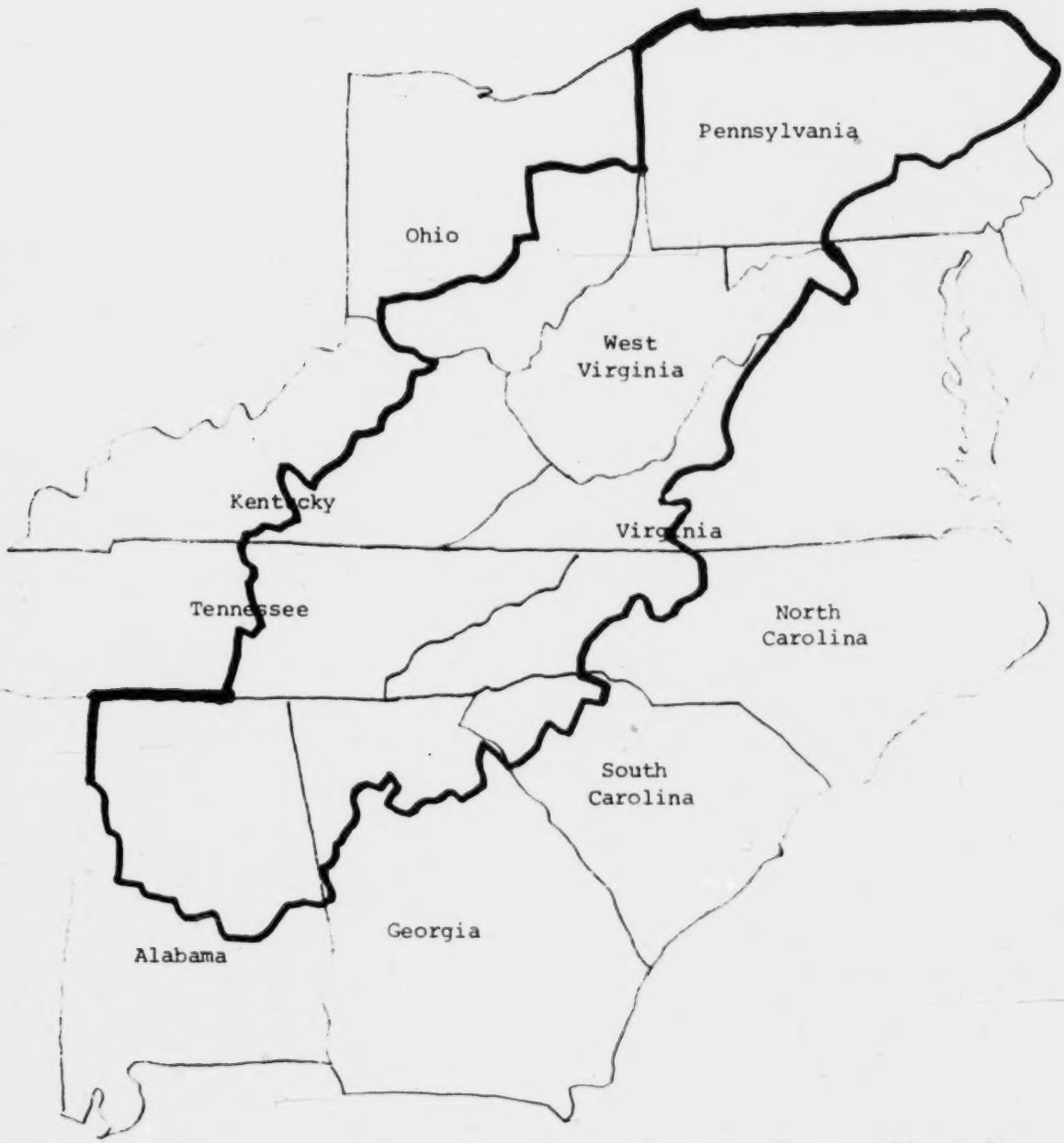


Figure 2

The Appalachian Region (Defined by Steering Committee for Organizing AEL) (Steering Committee, 1965, p. 2)

Feasibility Analysis

The feasibility analysis for the intended educational cooperative program was highly dependent upon a study done by consultants from the Systems Development Corporation (Kent, Davis, LeBaron, 1967). The basic idea of the Systems Development consultants was that a paper model for educational cooperatives based on reviews of literature, expert opinion, and possible field test would be misleading and not worth its cost. They made the following remark about what they called an adaptive methodology for development of the educational cooperative (ibid, p. 29):

> "No one yet knows enough about education technology to be able, merely by reviewing past experience, to write specifications for an effective educational cooperative. What is actually needed, is a systematic program of planning, thought, tentative decision, trial, evaluation, and revision. In other words, the specifications are to be developed through actual experience with a complete, real, cooperative, which is actually brought into being in constantly evaluated stages. Specifications must come after. not before, the first cooperative."

To achieve the openness of the adaptive methodology to penetrate into the unknown and to increase chances of success in approaching this neomobilistic decision setting, the Laboratory entered into arrangements with four field organizations. These field cooperatives were observed over a several year period.

Program Planning

The legitimacy of the needs identified for the Appalachian region and the apparent feasibility of designing and diffusing educational cooperatives gleaned from the feasibility adaptive methodology led to the statement of objectives for Appalachian educational cooperatives. The objectives formulated were as follows:

1. To make available for the participating districts cost effective educational programs and services on a regional basis.

2. To serve as a model of administrative practices which will enable participating districts:
   a. To analyze educational problems and devise solutions in an orderly, rational manner.
   b. To reallocate resources in order to achieve desirable educational outcomes.
3. To bring resources of other organizations (particularly state departments of education and institutions of higher education) to bear upon the problems of participating districts.

Consistently, with these objectives, a number of specifications pertaining to membership, governance, financing, and services were formulated. The program strategy called for the development of manuals dealing with the descriptive design, the process, and the structure of the educational cooperative. The descriptive design (AEL, 1971) provided a rationale for the educational cooperative emphasizing its character as a wild organization. The process manuals were to elaborate on the functions of the educational cooperatives as identified by Terry Eidell (Eidell, 1965). See Figure 3. The process manuals were identified as needs assessment, planning, programming, and evaluation. Structure manuals were identified pertaining to policymaking, information systems, personnel management, and business management. In all, nine documents were to be developed as a result of the program planning strategy.

Product Development Strategy

The product development strategy includes three stages of the Model for Educational Development. The stages related to product development very strongly are Design and Engineering, Field Testing, and Operational Testing.

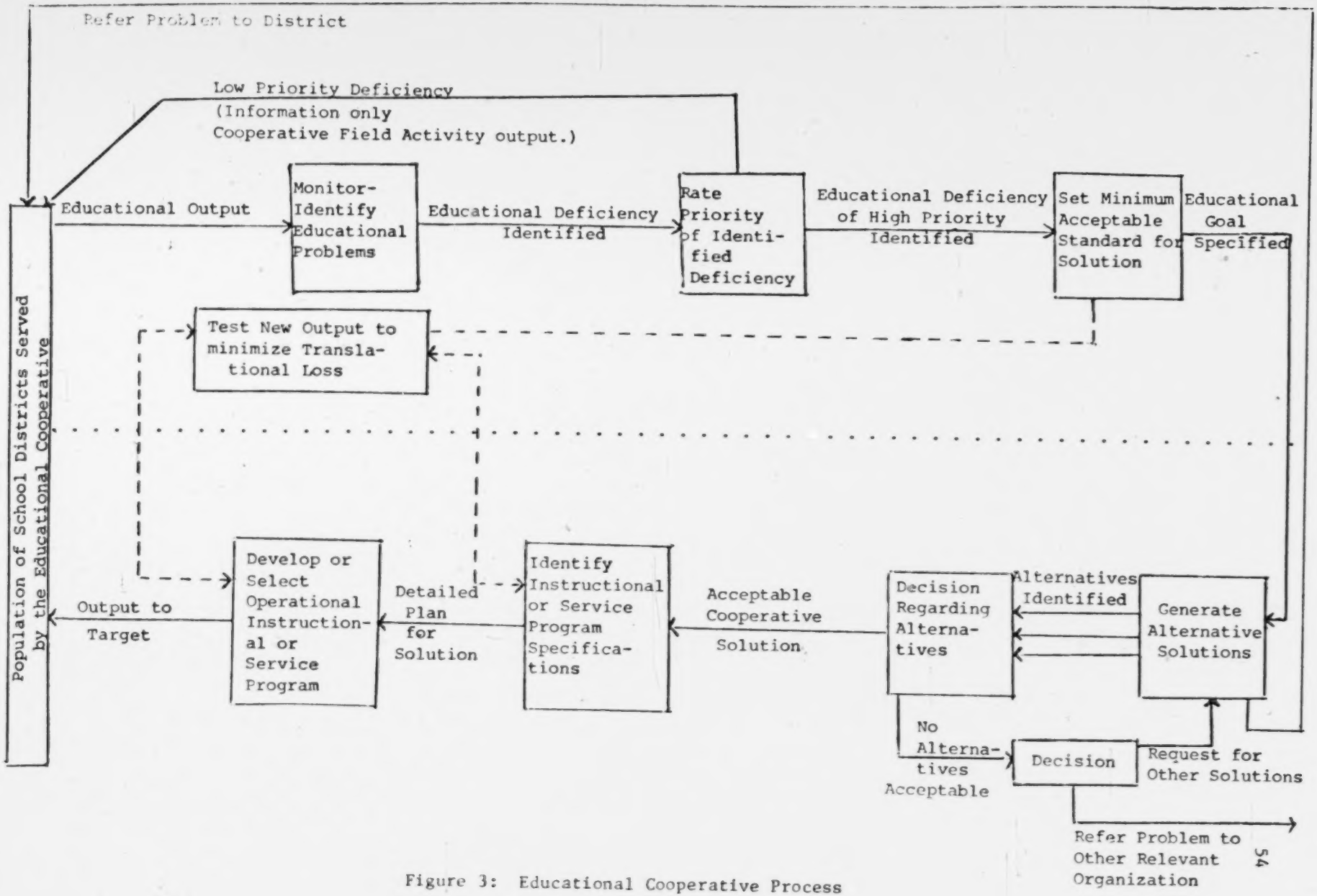


Figure 3: Educational Cooperative Process

The educational cooperative program was terminated during the Field Testing period. The discussion in this section will pertain to Design and Engineering, and the pertinent aspects of Field Testing with some extra Appalachian operational testing.

The product ensemble called for by the program planning strategy included nine manuals dealing with descriptive design, process, and structure of the educational cooperative. The descriptive design presented a rationale for the educational cooperative and stamped its character as a wild organization. The process manuals were to be developed consistently with the objectives and specifications for the educational cooperative derived during the program planning strategy phase of development. The process manuals were to cover needs assessment, planning, programming, and evaluation. The structure manuals were to cover policymaking, information systems, business management, and personnel management.

Criterion variables. Criterion properties are those characteristics of the thing with sufficient import to be used for judgmental anchors of satisfaction. A criterion model of a system is an idealization of the world to achieve understanding and control of that world. The elaboration of a criterion model involves the identification and designation of variety as the bearers of relevant information of social significance. The manuals were in themselves to be a model of an elaborated educational cooperative. The Educational Cooperatives were to realize their design, process, and structure from the use of the manuals. So the criterion properties of the manuals and the cooperatives were to be isomorphic. The criterion variety for the cooperative system was to have sufficient specificity to provide for the inclusion of requisite variety (Figure 4).

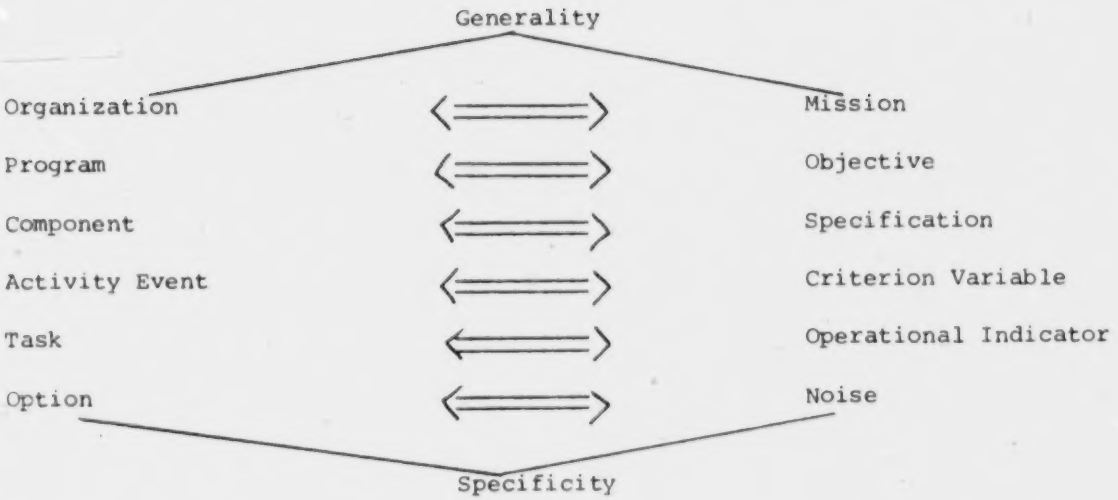


Figure 4

Organizational Design Correspondences
Generality-Specificity Continuum

An example of a criterion property of the educational cooperative provided for in the descriptive design is the property of inclusion. Any school district which can be benefited satisfactorily through membership in the educational cooperative is to be included. Other examples of criterion variables are those of priority setting, standard setting, and alternative generation.

Indicators. The criterion of inclusion was associated with an ensemble of operational indicators, viz contiguity, unitary intersection, contractual complement, conditions, radius, and enrollment. The determination of radius, for example, was to be achieved by use of scaled maps of the cooperative area to determine whether the cooperative service area is less than or equal to one hour driving time from the central location of the cooperative offices. Indicators as a rule are criterion-referenced manual-specific in content. Manual-specific operationality for methodological choices in setting priorities include synectic techniques, deductive techniques, inductive techniques, advocacy approach, and a mixed bag of other things such as sensitivity analysis, contingency analysis, afortiori analysis, etc.

Measurement and Instrumentation. Instrumentation is the invention or adaptation and utilization of devices to enable delineated information to be gathered for evaluative analysis. Instruments must provide for content validity, have parallel forms whenever possible, and have concurrent measures if practical. The instrumentation for the evaluation of the educational cooperative was conceived to gather information adequate for product decisions on the manuals. This required the consideration of mundane existence, multiplexing variety, operationality, and formative revision.

TABLE 1

INCLUSION INDICATORS

Indicator	Operation
Contiguity	Identify the district boundaries of members of the Cooperative on an official map of the region, and note the lack of disjointed members not sharing a border with another member.
Unitary Intersection	Note the mapped intersection of the Cooperative area and the adjoint planning and development district, and identify and enumerate those Cooperative members in the intersection.
Contractual Complement	Identify and enumerate any Cooperative members not contained in the mapped intersection of the Cooperative area and the adjoint planning and development district.
Conditions	Identify any conditions of membership imposed by the Cooperative.
Radius	Scaled maps of the Cooperative area shall be used, with speed zone data from the state highway agency, to determine whether the Cooperative service area is less than or equal to one-hour driving time.
Enrollment	The Cooperative's pupil population will be determined from school district data to determine if that population is between specified limits.

NOTE: Inclusion - to include in a given Educational Cooperative any school districts which can be served with satisfaction.

The first pass in the evaluation effort was conceived as the confirmation of existence of properties in the realization of the field cooperatives. In other words, the tone of neomobilistic plan change, one octave higher than mundane research, satifices with the existence of events and activities included in design. Initial efforts to attain an evaluative capability for the educational cooperative program defined a minimal basis for satisfaction with elaborated realization as the embodiment of design. The satisfaction gauges were the criterion variables. The criterion variables are comparable to factors in the multivariate space. Each criterion variable is an information channel multiplexed by the variety of a subset of the multivariate space. Factor analytic techniques commence with variety space and delineate the multiplex channels of the factors. The criterion variables are the multiplexed channels of satisfaction and the approach is the reverse of the factor analytic technique: an expanded variety subspace is sought out for the criterion variables to generate a variety universe for authentication of realization.

Twelve instruments were designed to get the information on mundane existence. These instruments covered separately the manuals for the process and the structure for the educational cooperative. Also, instruments were designed to get information on interaction of the board of directors, the minutes of the board of directors, the effectiveness of the educational cooperative, and an observational instrument was designed on the basis of an organizational taxonomic unit.

Evaluation plan. The plan for the evaluation of the educational cooperative was a general systems plan. The educational development was conceptualized through a state-space representation of institutional elaboration. The overall evaluation is fundamentally concerned with the

performance index of the institution; an adaptive convergence policy in conjunction with existing institutional states maps into a realization with some loss relative to design. The performance index is a weighted composite of terminal error, instantaneous error, and cost of control (Stepp, 1972).

A conceptual model of educational development may be based on self-organizing cybernetic systems. Predicated in the self-organizing system is the elaboration of realization in accord with an adaptive policy of convergence toward a designed model-reference. Continuing scrutiny of the need for the adaptation of the policy of convergent elaboration of realization is a function of formative evaluation. The model of elaborated realization as an embodiment of design is a function of summative evaluation. The institutional state-space, the formative convergence policy, and the realization of institutional elaboration should be relatable by mapping functions. The entire model is that of institutional automata. The educational cooperative in effect becomes a generator of a new range of decidability in the problem-solving coping behavior of school systems. The formative-summative contrast in the plan is exemplified in Table 2.

Evaluation Advisory Committee. An advisory committee was impaneled for the Design and Engineering phase of the evaluation. The panel was composed of three experts from state department, research administration, and university research communities. The principal recommendation of the advisory committee was the facilitation of communications between the product development staff and the evaluation staff to achieve a consensual validation of efforts.

Field Testing

The four field sites originally contracted with by the Laboratory for

TABLE 2

FORMATIVE-SUMMATIVE CONTRAST

Aspect	Emphasis	
	Formative	Summative
Aim	Tactics, specifications	Strategy, objectives
Emphasis	Technical efficiency	Theoretical adequacy
Scope	Fractional, infra-structure, micro-processes	Holistic, supra-structure macro-processes
Locus	Internal	External
Mode of operation	Negative feedback	Positive feedback
Decision points	Real-time	Discrete-time
User	Producer (Consumer)	Consumer (Producer)

the purposes of a feasibility analysis and field development of concepts about the cooperative were terminated as official organizations affiliated with the Appalachia Educational Laboratory. However, the organizations did continue as operational cooperatives. Four field activities were invited to apply to become the Laboratory's official continuing field test of the educational cooperative. One of these, the Upper Kentucky Valley Educational Cooperative, was chosen using criteria developed by the Laboratory's Educational Cooperative development team. A number of cooperative-type organizations, either newborn or in the process of being created, were invited to apply for consideration as a new field test educational cooperative. The Virginia Appalachia Educational Cooperative in southwestern Virginia was chosen as the new field test.

Evaluation Advisory Committee. In the Field Testing Stage, goal revision, reconsideration of criteria variables, revision of indicators, revision of measurement procedures and instrumentation, and revision of evaluation plan were achieved, and the Evaluation Advisory Committee review was conducted in the field testing phase of the educational cooperative development effort. The concerns of the Advisory Committee became focused on instrumentation and measurement techniques. The measurement and instrumentation approach had been toward confirmation of the existence of specified activities in manuals. There was some concern that existence confirmation alone was insufficient, which was undoubtedly so.

Visitation. Visitation was arranged with each school district in each cooperative area. A specially scheduled and conducted discussion pertained to each aspect of the process and structure for which a specific manual had been planned. Also each superintendent of each system was interviewed with

a special focus on the determinants of effectiveness of the cooperative in attaining its objectives as seen from the vantage point of the superintendent in his immediate district.

Base-line data. The base line data included documentary evidence of performance of the systems serviced by the cooperatives in those areas of practice addressed by the various product manuals. A number of bits of evidence were requested from the different school systems involved, including such things as salary schedules, planning documents, evaluation documents, and a selective multiyear crosswalk on fiscal program affairs. Before all base line data were procured, NIE called for the phase-out of the program. At this point all real progress in the program came to a halt.

Operational Testing

The relinguishing of control over field implementation of the design process and structure of the cooperative would have ensued official field testing. The degree of control in operational testing would have been much less in comparison to that in the field test itself. In the operational testing phase, the finalization of standards would have been made more public so that potential users could make a comparison of results from the operation. During field testing the standards would have been more flexible and subject to change in the formative upgrading. However, this was circumvented by the action of NIE and their imperative for phase-out of the Educational Cooperative Program.

Discussion and Summary

The construction of an evaluation model for a regional educational service agency, namely an educational cooperative, has been undertaken in

product development setting. The general evaluation design has been completed, but not validated. The construction of the specific design for the evaluation of the Cooperative was achieved during a more general emergent model-building effort for the evaluation of educational products (Bertram, 1973).

Basic issues inhere in the delineation, obtaining, and providing of information in a neomobilistic setting for planned change decisions. One very fundamental issue pertains to requisite variety for general systems decisions in educational development (Stepp, 1973). Any subsetting of the domain of educational development with system states, state variables, and state succession should have theoretical and methodological adequacy. The evaluation model must generate a systems decidability in terms of a basis for information, including incompleteness, consistency, controllability, and observability. Thereby, evaluation modeling may achieve the status of algorithmic advocacy in a satisfactory fusion of scientific rigor with institutional mission, priorities, and policy (Schutz, 1973; Stufflebeam, 1972).

BIBLIOGRAPHY

BIBLIOGRAPHY

Alper, Paul. "Introduction of Control Concepts in Educational Planning Models," Mathematical Models in Educational Planning. Paris: Organization for Economic Cooperation and Development, 1967.

Andrew, Gary M. and Donald E. Moir. Information - Decision Systems in Education. Itasca, Illinois: Peacock Publishers, 1970.

Anthony, Robert N. Planning and Control Systems: A Framework for Analysis. Boston: Harvard Business School, 1965.

Aoki, M. Optimizations of Stochastic Systems. New York: Academic Press, 1967.

Appalachia Educational Laboratory, Inc. Descriptive Design for the Educational Cooperative. Charleston, West Virginia: Appalachia Educational Laboratory, Inc., 1971.

_____. Educational Cooperative: Basic Program Plan. Charleston, West Virginia: Appalachia Educational Laboratory, Inc., 1972.

_____. A Strategy for Evaluation of the Educational Cooperative. Charleston, West Virginia: Appalachia Educational Laboratory, Inc., 1970.

Apter, Michael. Cybernetics and Development. New York: Pergamon Press, 1968.

Archibald, R. and R. Villoria. Network Based Management Systems. New York: John Wiley and Sons, 1967.

Armitage, Peter M., Cyril S. Smith, and Paul Alper. Decision Models for Educational Planning. London: Penguin, 1969.

_____. Models for Educational Decision Making. London: Penguin Book Company, 1969.

Ashby, W. R. An Introduction to Cybernetics. New York: John Wiley and Sons, 1959.

Beer, Stafford. Cybernetics and Management. New York: John Wiley and Sons, 1966.

_____. Decision and Control. London: John Wiley and Sons, 1966.

Bellman, Richard. Adaptive Control Processes. Princeton, New Jersey: Princeton University Press, 1961.

Berrien, Kenneth F. General and Social Systems. New Brunswick: Rutgers University Press, 1968

Bertalanfly, Ludwin von. General System Theory. New York: George Brayiller, Inc., 1968.

_____. "General System Theory - A Critical Review," Walter Buckley (ed.), Modern Systems Research for the Behavioral Scientists. Chicago: Aldine Publishing Company, 1968.

Bertram, Charles L. The Application of a Model for the Evaluation of Educational Products. Charleston, W. Va.: Appalachia Educational Laboratory, Inc., 1973

Booth, Taylor L. Sequential Machines and Automata Theory. New York: John Wiley and Sons, Inc., 1967.

Braybrooke, D. and C. E. Lindblom. A Strategy of Decision. New York: Free Press, 1963.

Bross, Irwin D. F. Design for Decision. New York: Free Press, 1953.

Broudy, Harry. Criteria for the Theoretical Adequacy of Conceptual Framework of Planned Educational Change. Bethesda, Maryland: ERIC, 1965. (ED 010 914)

Browder, Jr., Lesley H. Emerging Patterns of Administrative Accountability. Berkeley, California: McCutchan Publishing Corporation, 1971.

Buckley, Walter (ed.). Modern Systems Research for the Behavioral Scientist. Chicago: Aldine Publishing Company, 1968.

_____. Sociology and Modern Systems Theory. Englewood Cliffs, New Jersey: Prentice-Hall, Inc., 1967.

Campbell, Donald T. Exploration of Novel Research Designs and Measurement Techniques. Evanston, Illinois: Northwestern University, 1965. (ED 033 389)

Campbell, Merrill G. et al. Directions for Educational Development in Appalachia. Charleston, West Virginia: Appalachia Educational Laboratory, Inc., 1971.

Carlson, Richard O. School Superintendents: Careers and Performance. Columbus: Charles E. Merrill Publishing Company, 1972.

Center for Effecting Educational Change. Evaluation and "PACE": A Study of Procedures and Effectiveness of Evaluation Sections in Approved PACE Projects with Recommendations for Improvement. Report #1 of the second National Study of PACE. Fairfax, Virginia: Center for Effecting Educational Change, 1968.

Clark, David L. and Egon G. Guba. "A Re-examination of a Test of the 'Research and Development Model' of Change," Educational Administration Quarterly, 7(3):93-103, Autumn, 1972.

Cook, Desmond L. Management Control Theory as a Context for Educational Evaluation. Columbus, Ohio: The Ohio State University Educational Program Management Center, 1970.

Cook, Desmond L. "Management Control Theory as a Context for Educational Evaluation," Journal of Research and Development in Education, 3(4):13-26, Summer, 1970.

Croft, Don B. and Adolph K. Koenig (editors). Application of Research Management Techniques. Las Cruces, New Mexico: Special Interest Group on Research Management, New Mexico State University, 1971.

Curtis, Dr. William H. Educational Resources Management System. Chicago: Research Corporation, 1971.

Darlington, Richard B. and Jean F. Rom. "Assessing the Importance of Independent Variables in Nonlinear Causal Laws," American Educational Research Journal, 9(3):449-462, Summer, 1972.

Dubin, Robert. Theory Building. New York: The Free Press, 1969.

Eidell, Terry L. Model for the Management and Operations Sub-System of the Educational Cooperative. Eugene, Oregon: Center for the Advanced Study of Educational Administration, 1967.

Elgerd, Olle J. Control Systems Theory. New York: McGraw-Hill Book Company, 1967.

Etzioni, Amitai. Modern Organizations. Englewood Cliffs, New Jersey: Prentice-Hall, Inc., 1964.

Evans, J. A. A Framework for the Evolutionary Development of an Executive Information System, Part 1, Organizational Problem-Finding. Bedford, Massachusetts: Mitre Corporation, 1970. (ED 047 729)

_____. A Framework for the Evolutionary Development of an Executive Information System, Part 2: System Design, Implementation, and Evolution. Bedford, Massachusetts: Mitre Corporation, 1970. (ED 047 730)

Fels Institute. General Design for and Educational Planning-Programming-Budgeting System. Pennsylvania University, Philadelphia, Government Studies Center, 1968. (ED 037 812)

Finn, James D. "Institutionalization of Evaluation," Evaluation and "PACE". Fairfax, Virginia: Center for Effecting Educational Change, 1968.

Flament, C. Applications of Graph Theory to Group Structure. Englewood Cliffs, New Jersey: Prentice-Hall, Inc., 1963.

Gephart, William J. Criteria for Methodological Adequacy for Research on Educational Change. Bethesda, Maryland: ERIC, 1965. (ED 011 146)

Gephart, William J. The Problem and Problem Delineation Techniques. A paper presented at the Second National Symposium for Professors of Educational Research, Boulder, Colorado, November 2, 1968.

Gephart, William J. The Eight General Research Methodologies: A Facet Analysis of the Research Process. Bloomington, Indiana: Phi Delta Kappa, 1969. (ED 052 237)

Gideonse, Hendrik D. Educational Research and Development in the United States. Washington, D. C.: National Center for Educational Research and Development, U.S. Government Printing Office, 1970.

Gordon, Robert. "Optimum Component Redundancy for Maximum System Reliability," Operations Research, 5:229-243, 1957.

Griffiths, Daniel (ed.). Developing Taxonomies of Organizational Behavior in Education - Final Report. New York: New York University, April, 1968. (ED 021 339)

_____. Developing Taxonomies of Organizational Behavior in Educational Administration. Chicago: Rand McNally and Company, 1969.

Guba, Egon G. Methodological Strategies for Educational Change. Bethesda, Maryland: ERIC, 1965. (ED 011 404)

Habenstein, Robert W. (ed.). Pathways to Data: Field Methods for Studying Ongoing Social Organizations. Chicago: Aldine Publishing Company, 1970.

Haberstroh, Chadwick J. "Organizational Design and Systems Analysis," James G. March (ed.), Handbook of Organizations. Chicago: Rand McNally and Company, 1965.

Haggart, Sue A. (ed.). Program Budgeting for School District Planning. Englewood Cliffs, New Jersey: Educational Technology Publications, 1972.

Halpin, Andrew W. Administrative Theory in Education. New York: The Macmillan Company, 1958.

_____. Theory and Research in Administration. New York: The Macmillan Company, 1966.

Harary, Frank. "Graph Theory and Group Structure," R. Duncan Luce, Robert R. Bush, and Eugene Galanter (editors), Readings in Mathematical Psychology, Vol. II. New York: John Wiley and Sons, Inc., 1965.

Hawkridge, David G. and Peggie L. Campeau. Developing a Guide for Authors of Evaluation Reports of Educational Programs, Final Report. Palo Alto, California: American Institutes for Research in the Behavioral Sciences, 1969. (ED 043 677)

Hills, R. Jean. Toward a Science of Organization. Eugene, Oregon: University of Oregon Press, 1968.

Hughes, Larry W. et al. Interpretive Study of Research and Development Relative to Educational Cooperatives, Final Report. Knoxville, Tennessee: College of Education, The University of Tennessee, 1971.

Johnson, Richard A., Fremont E. Kast, and James E. Rosenzweig. The Theory and Management of Systems, Second Edition. New York: McGraw-Hill Book Company, 1967.

Jones, J. Christopher. Design Methods: Seeds of Human Futures. London: Wiley-Interscience, 1970.

Kaplan, Abraham. The Conduct of Inquiry. San Francisco: Chandler Publishing Company, 1964.

Keeney, M. G., H. E. Koenig, and R. Zemach. State-Space Models of Educational Institutions. East Lansing, Michigan: Division of Engineering Research, Michigan State University, 1967.

Kent, W. P., B. Davis, and W. A. LeBaron. A Preliminary System Description and Critique of the Appalachia Educational Cooperative. Falls Church, Virginia: System Development Corporation, 1967.

Klein, S. P., J. Burry, D. A. Churman, and M. A. Nadeau. Evaluation Workshop I: An Orientation. Monterey, California: GTB/McGraw-Hill, 1971.

Kohl, John and Charles M. Achilles. A Basic Planning and Evaluation Model for Cooperation in Providing Regional Education Services. University Park, Pennsylvania: Pennsylvania State University, 1970.

Lessinger, Leon M. and Ralph M. Tyler (editors). Accountability in Education. Belmont, California: Wadsworth Publishing Company, Inc., 1971.

Litterer, Joseph A. Organizations: Structure and Behavior, Vol. I. New York: John Wiley and Sons, Inc., 1969.

_____. Organizations: Systems, Control and Adaptation, Vol. II. New York: John Wiley and Sons, Inc., 1969.

Lutz, Frank W. and Laurence Iannaccone. Understanding Educational Organizations: A Field Study Approach. Columbus, Ohio: Charles E. Merrill Publishing Company, 1969.

March, James G. (ed.). Handbook of Organizations. Chicago: Rand McNally and Company, 1965.

_____ and Herbert A. Simon. Organizations. New York: John Wiley and Sons, Inc., 1958.

McNamara, James F. "Mathematical Programming Models in Educational Planning, Review of Educational Research, 41(5):419-446, December, 1971.

Messarovic, M. D., D. Macko, and Y. Takahara. Theory of Hierarchical, Multilevel Systems. New York: Academic Press, 1970.

Miller, Donald R. Planning, Developing and Implementing Title III, ESEA Projects. Burlingame, California: OPERATION PEP, May, 1968. (ED 022 247)

Mishkin, Eli and Ludwig Braun. Adaptive Control System. New York: McGraw-Hill Book Company, Inc., 1961.

Mullins, Norman. Cooperative Information System. Charleston, West Virginia: Appalachia Educational Laboratory, Inc., 1972.

_____. The Educational Cooperative Process Part IV - Draft. Charleston, West Virginia: Appalachia Educational Laboratory, Inc., 1972.

Neyman, Clinton A., Jr. Evaluation of ESEA Title I Programs for the District of Columbia, 1967-68, Final Report. Washington, D. C.: George Washington University, 1969. (ED 049 319)

North Central Association. Evaluative Criteria for the Evaluation of Secondary Schools, Fourth Edition. Washington, D. C.: American Council on Education, 1969. (ED 034 312)

Ogata, Katsuhiko. State Space Analysis of Control Systems. Englewood Cliffs, New Jersey: Prentice-Hall, Inc., 1967.

Organization for Economic Cooperation and Development. Mathematical Models in Educational Planning. Paris: Organization for Economic Cooperation and Development, 1967.

_____. Methods and Statistical Needs for Educational Planning. Paris: Organization for Economic Cooperation and Development, 1967.

Piaget, Jean. Structuralism. New York: Harper and Row, 1970.

Pontryagin, L. S. et al. The Mathematical Theory of Optimal Processes. New York: John Wiley and Sons, Inc., 1962.

Price, James L. Organizational Effectiveness: An Inventory of Propositions. Homewood, Illinois: Richard D. Irwin, Inc., 1968.

Provus, Malcolm. Discrepancy Evaluation: For Educational Program Improvement and Assessment. Berkeley, California: McCutchan Publishing Corporation, 1971.

_____ and G. Edward Lundin. "Evaluation for Administrative Action," Journal of Research and Development in Education, 3(4):1-108, Summer, 1970.

Roberson, E. Wayne (ed.). Educational Accountability Through Evaluation. Englewood Cliffs, New Jersey: Educational Technology Publications, 1971.

Roe, A. An Adaptive Decision Structure for Educational Systems. Berkeley, California: University of California, 1963.

Rudwick, Beranrd H. Systems Analysis for Effective Planning: Principles and Cases. New York: John Wiley and Sons, Inc., 1966.

Sanders, James R. and Blaine R. Worthen. "A Descriptive Summary of Framework for Planning Evaluation Studies," SRIS Quarterly, 5(1):10-14, Spring, 1972.

Schutz, Richard E. "Advocacy and R & D in Education," Educational Researcher, 2(1):3, January, 1973.

Scriven, Michael et al. An Evaluation System for Regional Labs and R&D Centers. Washington, D. C.: USOE (NCERD), 1971.

Shannon, C. E. "A Mathematical Theory of Communication," Bell System Technical Journal, 27:379-423, 623-656, 1948.

_____ and W. Weaver. The Mathematical Theory of Communication. Urbana, Illinois: University of Illinois Press, 1949.

Stake, Robert E. "PACE Evaluation," Evaluation and "PACE". Fairfax, Virginia: Center for Effecting Educational Change, 1968.

Starr, Martin Kenneth. Product Design and Decision Theory. Englewood Cliffs, New Jersey: Prentice-Hall, Inc., 1963.

Steering Committee, Appalachia Educational Laboratory. Quality Education for Appalachia: A Prospectus Proposing the Establishment of a Regional Laboratory: Charleston, W. Va.: Steering Committee, 1965.

Stephens, E. Robert et al. The Multi-County Regional Educational Service Agency in Iowa, Summary Report. Iowa City: Iowa University, 1967.

Stepp, Ermel, Jr. Planning Educational Systems Control: A State-Space Model of Strategic Factors. Ann Arbor: University Microfilms, 1971.

_____. Educational Cooperative Structure and Process: A Bibliography. Charleston, West Virginia: Appalachia Educational Laboratory, Inc., 1972.

_____. Evaluation Instruments for the Educational Cooperative. Charleston, West Virginia: Appalachia Educational Laboratory, Inc., 1972.

_____. General Evaluation Design for the Educational Cooperative. Charleston: West Virginia: Appalachia Educational Laboratory, Inc., 1972.

_____. Requisite Variety for General Systems Decisions in Educational Development. A paper presented to the Special Interest Group on Research Management, an affiliate of the American Educational Research Association, New Orleans, Louisiana: February, 1973.

Stufflebeam, Daniel et al. Design for Evaluating R&D Institutions and Programs. Washington, D. C.: USOE (NCERD), 1971.

_____. Educational Evaluation and Decision Making. Itasca, Illinois: F. E. Peacock Publishers, Inc., 1971.

Taylor, Donald W. "Decision Making and Problem Solving," Handbook of Organizations, James G. March (editor). Chicago: Rand McNally and Company, 1965.

Temkin, Sanford. A Comprehensive Theory of Cost-Effectiveness. Philadelphia: Research for Better Schools, Inc., 1970 (ED 040 503)

Tracz, George S. "An Overview of Optimal Control Theory Applied to Educational Planning." Paper read at the American Educational Research Association, Los Angeles, California, 1969. (ED 030 189)

Tyler, Ralph W. (ed.). Educational Evaluation: New Roles, New Means. The Sixty-eighth Yearbook of the National Society for the Study of Education. Chicago: The National Society for the Study of Education, June, 1970.

USOE. Testing and Further Development of an Operational Model for the Evaluation of Alternative Title I (ESEA) Projects. Washington, D.C.: Office of Education (DHEW), 1968. (ED 032 757)

Von Neumann, John. "The General and Logical Theory of Automata," Cerebral Mechanisms in Behavior. New York: John Wiley and Sons, Inc., 1951.

_____. "Probabilistic Logics and the Synthesis of Reliable Organisms from Unreliable Components," Automata Studies. Princeton, New Jersey: Princeton University Press, 1956.

Vroon, John. Rating Priorities, Setting Standards, and Generating Alternatives. Charleston, West Virginia: Appalachia Educational Laboratory, Inc., 1972.

Webb, Eugene et al. Unobtrusive Measures: Non-reactive Research in the Social Sciences. Chicago: Rand McNally and Company, 1966.

Weiner, Norbert. Cybernetics: Communication and Control in the Animal and the Machine. New York: John Wiley and Sons, Inc., 1948.

_____. The Human Use of Human Beings: Cybernetics and Society. New York: Doubleday and Company, Inc., 1954.

White, Ernest R. A Comparative Analysis of the Administrative Structure and Operation of Educational Cooperatives. Charleston, West Virginia: Appalachia Educational Laboratory, Inc., 1968.

Wilcox, Hayes. Business Management Guidelines for an Educational Cooperative. Charleston, West Virginia: Appalachia Educational Laboratory, Inc., 1972.

_____. The Educational Process Part III - Draft. Charleston, West Virginia: Appalachia Educational Laboratory, Inc., 1972.

_____. Policy-Making Guidelines for an Educational Cooperative. Charleston, West Virginia: Appalachia Educational Laboratory, Inc., 1972.

Willits, Robert. Personnel Management in an Educational Cooperative. Charleston, West Virginia: Appalachia Educational Laboratory, Inc., 1972.

Worthen, B. R. and J. R. Sanders. Educational Evaluation: Theory and Practice. Worthington, Ohio: Charles A. Jones (in press).

Young, Stanley. Management: A Systems Analysis. Glenview, Illinois: Scott, Foresman and Company, 1966.

Zemach, Rita. A State Space Model for Resource Allocation in Higher Education. East Lansing: Division of Engineering Research, Michigan State University, 1967.

ED 080610

TM 003 129

SECTION V

THE APPLICATION OF THE EVALUATION MODEL TO A CAREER EDUCATION PRODUCT

Charles E. Kenoyer

Appalachia Educational Laboratory, Inc.
P. O. Box 1348
Charleston, West Virginia

February 27, 1973

SECTION V

The Application of the Evaluation Model to a Career Education Product

The general goal of the Career Decision Making Program, as stated in the Basic Program Plan, is "...To prepare guidance materials and management procedures which will provide a core structure of sequenced, coordinated experiences designed to assist students in developing an understanding of themselves and of the world of work and in developing the capability to make knowledgeable career decisions." The strategy for achieving this goal includes the development of materials and procedures matched to student level, i.e., to elementary, middle, and secondary school students. The materials and management procedures are to emphasize (1) media, (2) student-activated resource materials, (3) curriculum-based materials, and (4) a systems approach.

Information System

The information system consists of several subcomponents. One of these is VIEW (Vocational Information for Education and Work), a file of aperture cards. These cards have job characteristics punched into the first several columns so that they may be accessed by means of a card sorter or other data processing equipment. The microfilm in the aperture card may be used with the microfilm readers at the VIEW Centers to produce a hard copy for the user.

A second component is a file of occupational information based on the Dictionary of Occupational Titles (DOT). A prototype file was developed and a file plan was written to enable school personnel and student assistants to construct the same information at the school.

The Worker Trait Group (WTG) index of the DOT is being rewritten. This index of the DOT provides a means of accessing the information in the DOT beginning with traits associated with workers. The rewritten version is intended to match the language level of secondary students more closely than the original DOT language does.

The Occupational Group arrangement of occupational information is provided in the form of a table of contents in the DOT. This arrangement has been duplicated as another means of accessing occupational information, and has been supplemented by listing WTG codes along with occupational groupings. Coordination of these two kinds of information enables the interested user to enter the system with an occupational group of which he has some knowledge and identify worker trait groups compatible with it. From this point the user can identify other occupational groups that are compatible with the same worker traits.

A keysort index is being developed which will allow system users to access the information file rapidly on the basis of WTG as well as aptitude/ability data. Like the aperture cards of the VIEW system, the keysort index is a way of cross-indexing on several variables. Its utilization does not, however, require mechanical or electronic apparatus as the computer cards do.

The last component of the information system is not focused on the Information File. It is an occupational information index which provides access to a variety of information sources.

Elementary and Middle School Components

Due to a pressing need for career guidance for students who will be on the job market soon, it was decided that the secondary school component

should be developed first. Except for preliminary planning, the work of developing the elementary and middle school components has not been done. Consequently, little evaluation planning can be done for some time concerning these components. A general evaluation plan has been written. For the present, however, the program effort consists entirely of development and evaluation of the information system and the secondary school component.

Secondary School Component

Structure of the Product

The secondary school component consists of the information system described above and an instructional sequence of sixteen units. The three general areas of knowledge addressed by the units are the self, the world of work, and the relationship between them.

Knowledge of self is approached through self-exploration by means of a number of attitude, preference, and aptitude instruments. The treatment of self also has an affective emphasis, attempting to induce in students a feeling that they have a degree of control of their destinies.

Knowledge of the world of work involves the concepts of Worker Trait Groups, the data-people-things (DPT) classification scheme on which it is based, and the occupational groups. Thus the world-of-work material is focused on enabling the student to utilize the information system's structure efficiently. Games and simulation exercises are employed to give students a feel for the complex pattern of tasks necessary to get to a finished product, for the arbitrariness of the particular task assignments that make up a work position, and other dynamic properties of the world of work.

The relationship between self and the world of work pervades the instructional sequence. Materials dealing with the world of work are organized according to worker traits, and materials dealing with the self are placed in the context of the world of work. Beyond this inherent coordination of the self and the world of work, however, there is an emphasis on a decision-making strategy by which students can seek and utilize information for career decisions to select from the available jobs those best suited to themselves.

A list of the instructional units and their topics is attached as Appendix A.

Current Stage of Development

With the exception of a few subcomponents of the information system which are in the Field Test stage of the model, the program is in the Design and Engineering stage. The evaluation activities associated with this stage include (1) developing and writing a plan for evaluation to be carried out over all stages, (2) assisting program developers in specifying criterion variables and objectives, (3) locating or developing instrumentation to measure the attainment of objectives, (4) designing assessment procedures, (5) selecting test sites and sample of students, (6) collecting data on program impact, (7) analysis and interpretation of data, and (8) reporting evaluation analysis and interpretation to program developers to guide revision.

Since the subcomponents of the secondary school component and the information system component are not all at the same level of development, a more detailed description of evaluation activities completed must be addressed to specific subcomponents. The evaluation plan encompasses all subcomponents as well as all future phases of development, but execution

of the plan begins with assisting in the development of criterion variables and objectives. This step and those that follow can be carried out only when program development has progressed to the proper point.

Activities associated with development of the overall program have included (1) specifying a set of goals for the total program, (2) developing a set of criterion variables for those goals, (3) identifying a tentative set of indicators (instruments or other measures) for the criterion variables, (4) writing an evaluation plan (which included a general plan for evaluating total program impact), (5) review of the evaluation plan by an evaluation advisory panel, (6) revision of program objectives to make them more specific to the realistic scope of the program.

The instructional units are in various stages of Design and Engineering. The first five units, and perhaps the sixth, are due for preliminary testing before the end of the current academic year in June. Development of the remaining units has hardly begun. Criterion variables have not yet been specified for the latter units and probably will not be specified until summer. Consequently, all the subsequent steps in the evaluation model must be deferred while both program development effort and evaluation effort are focused on those units to be completed this academic year.

Criterion variables and objectives have been specified for the first four units. Objectives for Units I through III have been reviewed and subsequently refined. At the time of this writing, Unit IV objectives have not been revised since August, 1972, and will probably need further work before instruments can be constructed for the unit. Instruments have been constructed for Units I and II, and instrument construction has begun for Unit III.

Motivated by time pressures, the developers took the initiative in locating preliminary test sites. This evaluation of the C D-M Program deviates from the model in that design of assessment procedures, followed by test site selection was not the sequence that actually occurred.

Two groups of students were selected by development staff. One of these groups is an intact class of 35 high-school sophomores. The other is composed of 10 volunteers who participate in 90-minute sessions twice a week at the Laboratory's offices. Materials are first tried out with the in-laboratory group, then taken to the classroom group. The in-lab group is frequently questioned in considerable detail about their reactions to materials. Valuable feedback is obtained from the in-lab students, who are paid for their services, before the materials are presented to the high school students in the more typical situation. An exception to this order of presentation was Unit I, which had been presented at the high school before arrangements were made for the in-lab group.

Assessment procedures have not been designed so much as worked in around the schedule set by developers. Production of materials, particularly media, is typically achieved just in time for presentation to the students. Under these circumstances, a major problem of assessment turns out to be getting assessment scheduled into the class periods at the appropriate junctures, with sufficient time allotted so that students are not rushed. The deadline-meeting problem tends to interfere with the requisite advance planning for such scheduling.

A second difficulty arises out of insufficient lead time in specifying unit objectives. Instrumentation has to be developed on the basis of these objectives, and pretests can be given only if instrumentation is developed

before unit materials are presented. When program developers are working to produce materials in time for presentation, the situation takes on the character of a race between developer and evaluator. The assessment procedures have, not too surprisingly, been somewhat variable.

In the classroom group, Unit I was assessed by means of a posttest only. Both a pretest and a posttest were administered to the in-lab group for this unit, but this group had already been exposed to Unit II (Decision Making) before the introductory unit was presented.

Unit II effects were assessed by means of a posttest only, in both the in-lab group and the classroom group. This procedure was the result of time pressure rather than preference. For reasons to be discussed below, a pretest-posttest group would be desirable for this unit.

Only a posttest will be used for Unit III. A pretest could probably have been carried out for this unit if it had been considered a high priority to do so. The nature of the unit makes a pretest less important during the Design and Engineering stage than for Unit II, discussed above. Both pretests and posttests are planned for Units IV, V, and VI.

Additional assessment is planned to indicate attainment of overall program goals. As a consequence of meeting with the evaluation advisory panel, the program goals originally listed were considered for revision, but the current set of goals suggested and the criterion variables that follow from them have not been established as the final version.

A questionnaire has been designed to assess student reactions to the filmstrip-tape presentations, of which there are an average of two per instructional unit. The filmstrips, which were contracted out, have been evaluated by the program development staff and, in some cases, have gone through multiple revisions before being presented to students. The

purposes of the questionnaire are (1) to assess the acceptability of the media to students and (2) to obtain any suggestions students can provide for their improvement.

Data have been obtained with several of the instruments cited, and interpretation of the findings has been provided to program developers for their use in refining the program. Posttest-only results on Unit I (obtained with the classroom group) indicated a high proportion (.81) of failure on one item. After some discussion between the developer for that unit and the evaluator, it was concluded that more emphasis was needed on the concept represented in the item. The pretest on the same unit (administered to the in-lab group) indicated a very low level of difficulty for two other items at pretest. This was interpreted as evidence that most students had met the objectives tested by these items before the unit was presented--rather than, for example, evidence that the wording of the items 'gave away' the correct response.

Posttest results on Unit II (Decision Making) indicated less than acceptable performance by several students. At this writing, resolution has not yet been reached on interpretation of this finding, but the results have been communicated to the unit developer. No pretest data were obtained for Unit II--unfortunately, since the objectives of this unit are related to general problem-solving ability and posttest level may not reflect gains.

Unit III (World of Work), on the other hand, does not require a pretest-posttest evaluation at this stage. The content is so unlikely to be known beforehand that posttest scores can be safely interpreted as indicators of program gains. During the Field Test stage, when the evaluation is intended

to inform persons other than the developers, a pretest-posttest design seems more appropriate.

The information system calls for a different evaluation emphasis. Like any program component the information system components are intended to make a difference in program impact, but like other reference materials their impact is less easily isolated for observation. Because the information system functions as the undergirding information source, it is not feasible to have a control group that receives the instruction but is not exposed to the information system. Therefore a factorial design is not possible; but criterion variables may be identified for assessing the effectiveness of the information system.

The most obvious of these variables is degree of acceptance by faculty and students. Acceptance could be indicated by a simple questionnaire.

Amount of use of the system might be considered an indicator of acceptance, particularly for students who are not under pressure to retrieve information contained in it. For students who are receiving the instructional units as either a core or elective course, the interpretation of a record of use would be different, but such a record would still be an important variable for assessing the system. If, for example, students receiving the instructional units use the system more than other students do, the interdependent nature of the instruction and the system of information is supported.

The information file is constructed by indigenous personnel (e.g., faculty, counselors, students) rather than produced by the program. What is supplied is the file plan, a set of instructions for constructing the file. A criterion variable for product evaluation, therefore, is whether the file plan enables the adopter to assemble the information file without further

instruction. Further assessment of the file plan is possible through a questionnaire indicating the user's opinion of its clarity, comprehensiveness, etc. Thus both performance (actual construction of the specified file) and attitude (rating of the file plan) can be used to evaluate this subcomponent.

The file plan has been subjected to preliminary testing. Secretaries of the Laboratory were able to construct the information file by following the instructions provided.

Subsequently the file plan was field tested at two sites, but in a version somewhat different from what is intended as its final form. The version that was field tested utilized the Ohio Vocational Interest Survey to generate student trait variables for entering the file. The final form of the file will be designed for other instrumentation. Field testing will then be necessary for the final version.

Discrepancies Between Model and Process

The evaluation of the Career Decision-Making program is a more recent undertaking than the evaluations of the other programs discussed in this symposium, and the product is a relatively straightforward one, consisting of an instructional sequence and resource materials. The model was therefore available and fairly simple to apply to the kind of evaluation required. There have been several mismatches between the model and what has actually been done, however.

The first discrepancy was that preliminary test sites and samples were not selected for a comparison group design as indicated in the model for collecting data in Stage 4 (Design and Engineering). However, since there are no competing programs in the area where preliminary testing is being done, the only appropriate comparison group would be a no-treatment group.

Nearly the same information can therefore be obtained through a pretest-posttest design, and obtained more easily. Consequently, the pretest-posttest approach has been followed in this program except where practical considerations prohibited, as described above.

A second discrepancy is the return to the refinement of product goals (or at least to development of criterion variables) after the step "Conduct Evaluation Advisory Committee Review". No branch back from the advisory committee step is shown in the model. In keeping with the model, however, preliminary testing (Collect Base Data and Collect Post Treatment Data, in the model) has proceeded on the early instructional units.

This recycling (or, perhaps more appropriately, retrogression) of overall program development while units continue through a normal sequence points to what may be considered a deficiency in the model: no distinction is made between end goals and subgoals. For programs that have an identifiable structure of subcomponents at some sufficiently early point in development, it would be useful to plan for evaluation of goals and subgoals on somewhat different schedules. Perhaps the interrelationship between goals and subgoals in a given program is such that goals should be determined first and subgoals derived from them, or the goals and subgoals may need to be refined iteratively, using the current state of each to guide revision of the other. The appropriateness of proceeding with unit development while overall program goals have yet to be settled upon is a question that could be addressed through such a structure a priori, rather than after the fact.

Another aspect of subcomponent evaluation that merits mention is that the experimental design (i.e., pretest-posttest, treatment-no treatment) approach is not applicable to every subcomponent to be developed. An example described above is the information system. Equally relevant are the

tape-filmstrip sequences for the units. Such subcomponents may be evaluated by means of expert opinion, explicit technical standards, student opinion, etc. The question, of course, is whether such activities are properly subsumed under formative evaluation procedures executed by program developers or whether they should be included in the evaluation model.

Finally, there is the problem of allocation of staff responsibility. Within the organizational structure of the Appalachia Educational Laboratory the distinction between program developers and evaluators has considerable practical importance. In other organizations this may not be a distinction between people, but only a day-to-day or moment-to-moment role distinction. To the extent that it is a useful distinction in either sense, the model might benefit from a clearer graphic separation of the development and evaluation components of the system.

APPENDIX A

CAREER EXPLORATION AND DECISION-MAKING
UNIT GOALS

1. The goal of the Introductory Unit is to provide students with an introduction to the Career Decision-Making Program and to develop an awareness of, and a readiness for, career exploratory activities.

2. The goal of the Decision-Making Unit is to help students become aware of the kinds of decisions facing them, and to motivate them to learn and apply a rational and systematic decision-making strategy.

3. The goal of the World of Work Unit is to provide students with insight into the basic elements of work, basic worker skills, grouping arrangements of work, and career development patterns.

4. The goal of the Vocational Goals Unit is to assist students in identifying and understanding some of the bases for, and exploring some alternatives to their expressed interests, vocational goals and career plans.

5. The goal of the Vocational Interests Unit is to provide students with information about their measured interests, to assist them in comparing their measured and expressed interest, and to help them make career plans that are consistent with their interests.

6. The goal of the Work Situation and Adjustment Unit is to assist students in understanding the various types of work situations that exist and the adjustment of workers to these situations, and help them explore the variety of occupations related to them.

7. The goal of the Work Conditions and Physical Demands Unit is to familiarize students with working conditions such as the environmental and physical requirements of different careers so they can determine the appropriateness of various occupations.

8. The goal of the School Achievement and GED Unit is to assist students in understanding the relationship of their school achievement and general educational development to various occupational groups and levels within those groups so they can explore appropriate occupations related to that level.

9. The goal of the Aptitudes Unit is to provide students with information regarding specific aptitudes considered necessary for individuals to be successful in various occupations, and to provide students with information regarding their specific aptitudes so they can explore various occupations and establish goals that are commensurate with their abilities.

10. The goal of the Work and General Experiences Unit is to help students recall the important things they learned about themselves and the world of work from previous occupations and to consider their leisure time experiences that have relevance for career planning.

11. The goal of the Economic Influences Unit is to assist students in understanding how economic conditions and structure influence their present decisions as well as their future careers.

12. The goal of the Social and Family Influences Unit is to help students become aware of the pressures and expectations exerted by their culture, peers, and family and to assist them in asserting themselves so they can become the kind of person they want to be.

13. The goal of the Decision-Making and Career Planning Unit is to help students incorporate information and decisions made during the previous units and implement the decision-making strategy, through simulated experience, to make a knowledgeable educational decision and begin career planning.

14. The goal of the Educational Planning Unit is to assist students in exploring possible educational programs so they can make an appropriate academic and/or training selection.

15. The goal of the Your Future Unit is to help students look into the future and develop a career plan of action that has built in periodic reviews to allow for career progressions and career alternatives.

SECTION VI

INTERACTION OF DISCUSSANTS, PANEL AND AUDIENCE

Gary Borich
Desmond L. Cook
Panel and Audience

SECTION VI

Interaction of Discussants, Panel and Audience

During the symposium, the Model for Evaluation of Educational Products was described, and three programs were used to illustrate certain aspects of the model. Following the discussion, the discussants gave their reactions and all participants became involved in an extended interaction with each other and with the audience. A transcript of the interaction follows, and has been slightly edited for clarity.

Dr. Hemphill

The first discussant will be Dr. Gary Borich, who is Director of Evaluation at the Texas R and D Center.

Dr. Borich

I have several questions to raise. I'm always surprised when I hear of a new evaluation model. It seems there are more models than evaluators. This is particularly distressing because each model purports to be a universal answer to all our evaluation problems and in reality it is usually applicable to a very small subset of evaluation problems. I think that with all the models and methodologies it is time for evaluators to stop generating models and to pull together the critical components of the models.

I have three basic questions. If we take the idea of a model literally, as a plan of action, the first question is "In what situation is that model applicable?" For example, here today AEL is suggesting a model. Is this model specific to a subset of problems that might be in the Appalachian region or is this something that someone might use no matter where they are? If the model is a specific one, applicable only in a certain number of

situations, what characteristics must a setting have for this model to be useful? Those characteristics should be specified as part of the model so it does not purport to be universally applicable in all situations if indeed it isn't.

The second question involves methodology. A great weakness of all models is that very few of them are specific about the methodology that they imply. They suggest plans for action but they do not explicate the specific techniques and procedures necessary to implement the model. If the model involves context, input, process, and product, surely then each of those areas must have some kind of methodology that an evaluator can use. Words alone and concepts alone are not enough to provide a sufficient evaluation model. It's the methodologies that need the development. The jargon, although helpful to the development of methodology, is superficial to the real question of how do you perform an evaluation and collect the data, and what kind of statistical methodologies are applicable to what kinds of data. So my question is "Do you address the issue of formative and summative evaluation in the model? To what extent does the model imply various methodologies, and if it does imply these, what are the methodologies?"

The third question involves the issue of dissemination. Evaluators are sophisticated enough, at this time, to realize that there are a lot of different kinds of information coming out of evaluation models. The information is not always concerned with decision-making. It's not always concerned with a heuristic for research. There are different kinds of information for different people. When you're talking about a complex model, the implication is that at certain points in time there are a lot of <u>different</u> kinds of information that should come out, and the AEL model

does bring up the point of different kinds of information. My question here is "What kinds of information can you derive from this model that might be of value to different people? More importantly, to whom would various evaluation reports be geared?" That is, what is the level of specificity for various reports that could be derived from the model and to whom are those various reports going?

Those are my three questions. Maybe some additional ones will arise in the course of the discussion and hopefully they will provoke interaction from those in the audience.

Dr. Hemphill

We will keep track of these questions and come back to them after Des has had an opportunity to react. Our next discussant is Dr. Desmond Cook, Director of the Educational Program Management Center at the Ohio State University.

Dr. Cook

The nice thing about being the second discussant is that you can either reinforce what the first one has said or agree with what he has said. I did agree with the several points Gary was making.

One of the problems is that one cannot do justice to large amounts of work, a lot of effort and energy put into developing a model in about five or ten minutes as a discussant. All you can do is kind of come at it with broad brush strokes and say these are some things that are of concern. I don't want to nitpick on semantics and this kind of thing. I think it's rather important, however, that you at least know the position I'm coming from when I look at it. I'm coming at it from what might best be called today systems management theory or management theory or systems theory. Therefore I have that perspective when I look at things.

When I see words like "A Model for Product Evaluation", I begin to get hung up immediately on semantics. We've already discussed a little bit of "What do you mean by model?", "What's the function of a model?", "What's a product?", "What really is evaluation?" For example, I can think of products at many levels. A product to me could be the knowledge we gain from the moon landings and its impact upon society. Do you evaluate that product the same way as you would evaluate the product of actually landing on the moon itself and all the attendant activities thereto? I can also think of a product being at a capsule level; just the capsule itself, not the booster or anything else. That's a product. Are we talking, when we use the term "product", about all three of those? Are there levels we have to deal with here? Are we dealing with terminal products or packages or are we dealing with interim products? This is the point that I believe was raised by one of the speakers here. To me there is a lot of reflection that this model refers to a terminal product, it's an end package. In my own thinking there are a lot of products being developed as you go along the way, call these sub-objectives or whatever you want. I think the question then becomes "Does the model apply to those products as they are being developed as well as to the overall package or the terminal product at the end?" It appears right now as I read the papers that it deals with the terminal product.

I was looking for criteria which would tell me if I had a good product. This point comes with a slightly hidden agenda because about a week after I accepted the invitation to appear on the panel I ran across a paper in the journal Research and Development titled "Six Keys to Evaluate New Consumer Products". It stated that "This article describes a model to assist engineers to evaluate new and unique products. The basic model

is applicable to a variety of consumer products. Its principle advantage is in providing a common, systematic conceptual framework by which to measure key qualities influencing acceptance of consumer products." Then they go on to list the six criteria, things like safety, durability, efficiency, effectiveness, asthetics, economics, special features, etc., and give you a case history of how this model can be applied. I was struck by the similarity. I think the point they were stressing was not the process that was involved, but rather the criteria by which the product could be assessed. That was one thing I didn't quite see in the paper that I was hoping to see. What are the criteria by which we judge educational products? Are there six or seven basic ones? This happens to be an industrial example, but they have identified at least six of their concerns.

What I saw the papers dealing with was a basic process which is largely described in what I want to call the systems design-systems development literature. Basically, we have a kind of problem-solving model when vou get into the fundamental roots of it. The process is fairly well-known and fairly well described in many different kinds of journals. This is not to negate the work which has been done. I would certainly hope we haven't gone through the process of reinventing a process with which we are already familiar. It talks in the paper about devising an ideal sequence. That's like devising an ideal wife. You'll never have one. You'll never find one. If you set up such a beautiful set of standards for an ideal evaluation system, you may wind up without any either. You can never find a reality that fits the ideal. If fact, that becomes a convenient way never to have one.

The model assumes a lot of rationality. My experience in living a few years is that a man behaves awfully irrational. You can present all

these nice flow charts and he isn't going to behave that way at all. All the model and models like this can do is give you some bases which you ought to touch as you move through this thing. And to make sure you have them covered whether it's done in the stated fashion or in some other order. The point is that these are the things which have to be considered. I think the model is effective for that.

I have some concern about product evaluation from whose viewpoint. There was some discussion of this. Are we looking at the product from the point of view of the user or the producer? I happen to have as kind of a hobby the flying of radio-controlled airplanes around the sky. The manufacturer may produce a beautiful transmitter-receiver combination which he thinks is ideal; it's gone through the process and so on but I might not like it as a potential consumer. It doesn't meet my needs. And so you always have to be thinking in this particular dimension.

Another question I dealt with was this: Are we describing a monitoring control subsystem or a larger production system. Is this basically what the model is? Is it a monitoring control from management's point of view? Discrepancies are noted between certain standards, the products look like they are recycled, etc. This is what you read in the management literature in control theory, cybernetics theory, and system theory. This idea has been incorporated and to me that's good. My question is "Is there any way it can affect the production system?" Can it begin to feed information to say this is the way the production system ought to be changed. Not only the product being produced but also the production system? And if it is, then it's got to be integrated in that kind of way."

Most product models deal primarily with what I want to call performance specifications; what should the product look like? how should it function? etc., but tend to ignore two very crucial variables commonly called schedule and cost. One of the real problems here is that these three things get integrated very tightly together and when you start shifting around performance specifications for a product you certainly begin to play with schedules and budgets. You've got to think of all three simultaneously, not just one. I don't know whether the model was designed to provide that kind of information or whether you have another system which provides somebody with schedule and cost information and a second system which provides somebody with performance specifications. I think somehow we should work to get the system to include all three because all three are very crucial and should be presented at the same time, particularly if you are going into recycling. What's the cost of recycling? What time do we have?

The papers ask for suggestions for alternatives to this particular model. The question that comes to me is "Are there any design criteria for these alternatives?" What can they be? What kind of design criteria can we have for product evaluation model? What are the criteria against which we can judge the present model? This is a product. How good is it? What are the criteria by which we judge how good this one is? Is it pragmatic tested? Does it work? Function? Are there design criteria by which we say this is a good model or a good product? Can you apply the model to this product itself?

If some of you are really interested in the problem of designing artificial things, and I'm not creating that idea, it comes out of Herb Simon's booklet called The Sciences of the Artificial in which he talks about how man puts together things which help him accomplish his purposes. It pre-

sents some excellent chapters on problems associated with design, how you create these artificial processes and functions, etc. to help you get your job done. And I certainly encourage you to read it because he gives some of the criteria and thinking and logic by which designs can be formulated.

Dr. Hemphill

I wonder to what extent that a model such as this is a consequence of a bigger organizational plan for a Laboratory? For example, it looks like we have in this organizational chart for the Laboratory a whole division concerned with evaluation. Would you have the same model if you made evaluation a function that you associated with the operators of each of the programs in the Laboratory? To what extent does the model depend upon grander organizational skills that govern the whole process?

Dr. Bertram

I think this model is very much a consequence of the organization of the Appalachia Educational Laboratory at which it was derived. Other R & D organizations have evaluators within programs as opposed to a separate evaluation staff, and it's been my observation that you find more emphasis on the in-program, formative-type evaluation whereas this AEL model appears to me more summatively oriented. Dr. Cook alluded to that. This model is concerned more with the end product-type evaluation. I think that it is very much an artifact of the laboratory staff organization, and one possible lesson is that you determine what kind of information you want and then decide how to organize to conduct the evaluation.

Gary alluded to "what kind of information do we aim at different audiences?" We do have different audiences, Gary. One is the university type with whom we must meet certain standards to communicate effectively. We do our statistical analyses of various kinds where appropriate. Brainard

alluded to some of them when he described the Early Childhood Education evaluation. At the same time we serve an audience of users, many of whom are not even interested in how we arrived at the information. They just want to know "what will the product do" and if the Lab will stand behind it. To elaborate just a little on the reporting system that Brainard was talking about, for a given year with the Early Childhood program we'll produce perhaps a dozen technical reports, one summary report of 30 pages, and then, for the key decision-makers, perhaps a one-page abstract of the "whole thing". We have a large volume of information but a decision may be partly based on a one-page sheet of paper as opposed to reports for technical quality. However, if we didn't have technical quality as backup, that would catch up with us, too.

Another question concerned the criteria by which we judge the effectiveness of an evaluation model. If the definition of evaluation involves decision-making, then, we should look at the different people who are making decisions and assess whether or not this information did enter into their decisions. In some way, we need to find out if the program directors are using evaluation information supplied by evaluation to make their day-by-day decisions. That would be one indicator of the effectiveness of evaluation. Another would be the use of the evaluation information by the clients of the Laboratory. A third one is NIE -- do the funding agencies use the information supplied by evaluators?

Dr. Kenoyer

Two other questions by Dr. Cook are closely related to that - one has to do with cybernetics and control aspect - if you can make the distinction between control and information, which you can in some contacts and not in others, I think what we attempt to do w th our evaluation is to inform

those people who are decision-makers. To the extent that they are rationally determining their decisions on the basis of input there is control or to the extent that they are being forced by someone else to pay attention to our evaluation there is control. But this is something that is more a total systems question than an evaluation-level question. One item of information that enters into this has to do with the schedule and cost question, particularly the schedule. If in fact we can do our evaluation in a context where attention is either forced or elicited in some way, then pressure is exerted on the developer not only to do the product on time but to do it right on time because he knows he is going to have to recycle (which will mess up his time schedule) if it doesn't meet the standards. So I think all of those questions sort of form a nexus and have to be treated together.

Dr. Hemphill

You talked about decision-makers, that you are doing evaluation to provide information for decision-makers. I'm sort of confused about who these decision-makers are that you're talking about. Sometimes it seems to me that they are the managers of the program, or the Laboratory director; other times it seems to me like it might be potential users, etc. You made a comment that the information didn't get available in time to help the decision-maker, that he had to make the decision without the information. Is there a question of having information that the decision-maker doesn't fit into his payoff matrix, which is irrelevant as far as he is concerned. How does this model take care of questions such as those?

Dr. Cook

You are considering a question which deals with what I want to call the behavioral dimensions of information systems, that is, how does the

decision-maker himself behave as a human being. Almost all of these models I keep running across assume a kind of rationality. The persons are coming at decisions in many different kinds of ways with many different viewpoints, political, social, and so on. And I think you could have to some extent a beautiful system which provides him a lot of information which he'll never use. For behavioral type reasons.

Dr. Hemphill

This seems to me a very good thing to separate very carefully. You talk about a model. If you've got a strictly rational model, go to a mathematical model, there is one such that's been around for a long time, statistical decision theory and I'm not sure to what extent your thinking is related to that rather abstract mathematical model. Or is it a behavioral model, that is, how do the decisions get made? It is fairly clear that mathematical, statistical decision theory is not very often used by a person who is making the decision. When you talk about model development, what kind of a model are you developing? One that is abstract, that in the extreme would be a mathematical model? Or is it a behavioral model?

Audience

It was alluded to very late in the discussion period, that was schedule and budget. And one of the things that I have never seen written in the area of education is the rank order of quality, budget, schedule, and some of the other kinds of things we do in the product development area. I would like someone to respond to that.

Dr. Cook

Let me respond to this. I have a graduate student this year, currently doing a dissertation on trying to ascertain the relative ranking of criteria for judging when a project is successful. We're looking at six possible

criteria, one of which happens to be called performance, that's meeting the specifications kind of thing. The second is schedule, the third is cost, the fourth is spin-off to the institutions. A fifth is customer satisfaction. The sixth is what we're calling follow-on work. That is, you did such a darn good job on this product that someone is coming banging on your door because they like what you did. We hypothesize that different organizational environments will have different criteria. I would believe, for example, that my dean would argue that my project is successful because there's spin-off in the sense that we have better staff and he doesn't care what the product really looks like. We've got more and better trained staff as a consequence. So this criterion of success of the project is more important to him than whether I stayed on schedule. So we're looking at these six criteria from a project director's point of view, a dean's point of view, and a sponsor's position. In most literature that I've read, and there's been a lot of research done, performance is the highest ranking criteria, i.e., making sure you meet the objectives that's the criteria that satisfies.

Audience

What are your criteria for judging your criteria?

Dr. Cook

There was an article written in Ed. Psych. about twenty years ago called "Criteria for Criteria".

Audience

It would seem to me in part that the discussion is about the question of money and scheduling and the answer being relevant to the person who uses or looks at that product. That person in the example is in higher education and institutions may be looked at by two people differently. Surely it is

obvious that a model for all the things that have been discussed as it is implemented in public schools needs to reflect that uniqueness of that individual concern, how it looks at it, its cost in relation to what it's going to buy, and along the way, having to look toward a funding agency like NIE. If they want that model to satisfy their needs, the Laboratory's needs, as a model for future funding, they'd better be cognizant of those criteria. On the other hand, if that model is needed to assess early childhood needs or where those students are for an individual program which is specific to that area, it should have that uniqueness before it's an applicable model. I think what's most important in just the last half hour of discussion is that anyone considering taking a broad model for use would be ill advised to say the least.

Dr. Kenoyer

In answer to the question of who the decision-maker is, the decision-maker is a different person at different times during the development of the product. In regards to this final comment you made, in trying to cover all the bases with one model, I think the best we can do, or the best we have tried to do here, is to cover those kinds of information (and sometimes this doesn't even appear in the chart) but to cover those kinds of information that are needed at specific stages for specific decision-makers. As long as we're talking about a program developer, as we are in the Design and Engineering Stage, essentially we're talking about formative feedback, we're talking about a close-order kind of interaction in terms of fine grain time span and we're also talking about diagnostic information as opposed to overall evaluation of how things look on the whole. We would like to know specifically whether each objective is being met. We would like to know if a given objective is not being met, precisely how are we deviating

from it? When we get to the stage of Field Testing, and especially the Operational Testing, we are reporting to the potential consumer. He couldn't care less what kinds of diagnostic things we found earlier, what he'd like to know is, how does the thing function now that it's finished? He also coundn't care less what the cost to develop it was. What he wants to know is "What is it going to cost to apply the thing?" We do in our technical reports on summative evaluation consider cost effectiveness as part of performance effectiveness.

Dr. Bertram

Let me respond to a comment by Gary. I did not mean to imply that we saw this model as a panacea for all educational evaluators. There are some activities suggested in the model that have application over a broad variety of settings. We do think it has some application to the products currently under development in our own Laboratory.

Dr. Borich

Let me follow up on a comment that I've been thinking about in the last few minutes. Des mentioned determining criteria for the success of a product. At the risk of making that paradigm even more complex, we already admit that criteria change depending upon whose criteria you're using. It seems those same criteria change depending upon what the product is. As the product changes, the various criteria will change and in turn the people who have criteria will change. The reason I brought this up is that in the paper it is suggested that it takes a year to go through this cycle of the model. It seems clear that this would be dependent upon the product and that any kind of universal statement that this model is applicable to all products and would take one year to implement would seem to limit its effectiveness. Clearly the model would have to be geared to the

kind of product in which you're engaged. This goes back to Des' comment "What kind of product is a research product?" Is it a hardware product? I take it that you didn't mean that the model would take a year to implement.

Dr. Bertram

No, I think that the suggestion was "perhaps a year" just to give some temporal indication of the development schedule for products in our own lab. But many products are on an annual cycle parallel to the annual school year. Some product elements have a field testing duration of only a few weeks. So the time does vary considerably with different products.

Audience

I see down here product implementation at the end of the model. What does this imply? If it implies adoption and institutionalization by the user, is the product abandoned at that stage to the user? Is there no need for some feedback as to what is happening to it?

Dr. Bertram

Right, that's a good question. May I respond to that? We would like to be able to follow the products developed by the Lab over a period of years. That's really the question, isn't it? How long do we intend to monitor the products to make sure they are continuing to meet the performance standards which were set earlier? Unfortunately, with the funding arrangement as it is now, a program's funding may be for five years and this funding is discontinued, so unless there's a strong organization within the Lab which can monitor the continued effectiveness of products, we don't have the resources to do it. Funding is usually discontinued following the development of a particular product We don't presently have the funding to do the follow-up study we would really like to do. We recognize that problem and I wish there

was some mechanism by which continued follow-up of developed products could be guaranteed.

Audience

Some monitoring is taken care of by new needs assessments being conducted. I assume that needs assessments would be sufficiently sensitive to the changes in products being used and so while that specific product may not be monitored on a one-to-one basis, I would hope that through a needs assessment, at least some of the impact could be judged.

Dr. Cook

We're developing a package for RBS and one of the questions NIE is asking RBS right now has to do with impact. Data on impact, have we made a difference with that package (in this case, the management of projects within the school district) and can we document it? If we can get that kind of impact data, and that's not just immediate, this is down the line, then they're more likely to sustain further work on the product. If not, they're simply going to abandon it. I think that's the question he's asking, and what your problem is, is getting that impact data a year or two years down the road to see whether you really made any difference or not. That's the kind of data I understand NIE is beginning to ask for, more and more. This puts you back in kind of a research mode, I guess.

Audience

My comment is more of a recommendation than a question. As I observe your model you display seven stages. And the label you are putting on those phases is very revealing. Under product planning, you are assuming the specification of outcomes. Am I correct? (Right.) If I revise this model, I would use that specification of outcomes as one of the major phases. Several of the questions have related to a concern about how you specify all

the dimensions you have to specify. That means you have three dimensions in which you should specify outcomes. One is the systems level, the benefit area, the social impact, impact on education, large context in which the product is to be used. The second level is obviously product specification, in this case, what should the user be able to do having used the product? This is typical, most often quoted in terms of objectives of the product. The third level refers to questions which relate to the utilization of the product, such as cost, maintenance problems, usability of the product, criticality of the product, and if you have examined all these three levels, then you will get a more detailed specification which will allow you, the evaluator, to ask questions of all these levels.

Dr. Kenoyer

I think the reason that we phase in rather strongly on Step 4 instead of earlier is that to delineate the evaluation activity in such a way that we are talking essentially from the standpoint of our organization which has an evaluation group. An evaluation group, per se, is not called into the planning phase although individual members of us might participate in it. So the model is to that extent shaped around the particular organizational structure we have.

Audience

The more important questions should be asked during those 1, 2, 3, stages...

Dr. Kenoyer

I think some of the questions are asked. It's just that we didn't see that as part of the evaluation activity.

Dr. Stepp

I think he's making a very important point, too. Though we have laid down the claim that this has been a descriptive model, the point now is, is

it going to be a prescriptive model? I think for it to be an adequate prescriptive model, it must give attention to these essential program strategies. If it does not, then the model will be inadequate.

Dr. Bertram

This brings up a more general question, too. What is the domain of the evaluator?

Audience

To aid in decision-making.

Dr. Bertram

To aid in decision-making. But when you get down to finer points, I'm not sure we have it. Since our time is gone, let me conclude with one final point.

To me this points to an even bigger need than just at AEL. I think evaluators at the different institutions working in Research and Development need to develop some kind of a consensus concerning what we are all doing that's acceptable. What are our standards for evaluation of products? They're a little different from the standards for research, which are fairly well worked out. We need input from different disciplines, such as systems management theory which Des represents. I would hope that this would be a continuing dialogue which will result in more consensus concerning the criteria by which the effectiveness of product evaluation is judged.

Dr. Hemphill

Adjournment